THE UNIVERSITY *of* EDINBURGH

THE OPERATION OF EYE-MOVEMENT CONTROL MECHANISMS DURING THE PERCEPTION OF NATURALISTIC SCENES

R. CALEN WALSHE

THE UNIVERSITY *of* EDINBURGH

Department of Psychology
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

R. Calen Walshe: *The operation of eye-movement control mechanisms during the perception of naturalistic scenes,*

ABSTRACT

Understanding of visual scenes takes place within very brief episodes known as fixations. To explore the extent of the scene, the eye shifts between fixation locations at intervals of roughly 300 ms. Currently, it is a matter of open inquiry as to what factors influence the timing of these movements. This thesis focuses on understanding the mechanisms that govern the rapid adjustment of fixation and saccade timings when novel stimulus information is encountered during a fixation. In part I, I use an experimental technique known as the fixation-contingent scene quality paradigm to control the quality of incoming visual scene information. This approach is used to assess how fixation timing adapts to moment-by-moment changes in the quality level of the stimulus. I find that quality changes tend to result in an increase in fixation durations and this occurs whether the quality is increased or decreased. Using distributional analytic techniques, I argue that these results reflect the combined influence of a rapid surprise related process and a slower acting encoding related influence. In part II, I study how fixation durations are influenced by the underlying saccade programming mechanisms. An important assumption within the eye-movement control literature is that there exists a threshold called the point-of-no-return. Once this point has been reached, a saccade may no longer be modified or cancelled. I adapt a classic psychophysical technique known as the double-step procedure to study the point-of-no-return within scene viewing tasks. I also provide a measurement of the saccadic dead time, the last point in time that a saccade may be modified. In Part III, a formal model of fixation durations in high-level tasks is presented. I build on recent modelling work and develop a formal account for the early-surprise late-encoding modulation account of fixation durations in scene viewing tasks. The model is tested against data observed in Part I of the thesis. I demonstrate that the model does a very good job of predicting these

distributions with relatively few assumptions. In summary, I use experimental techniques in combination with computational modelling to reveal how a composite of low-level (saccade programming) and high-level (information processing) considerations can, and must, be taken into consideration when understanding eye-movement control behaviour in scene viewing tasks.

PUBLICATIONS

Two published papers have been included as chapters in the thesis:

Walshe, R.C. & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision, 15(5)*, 1-19.
Appears in Chapter 1 within Part I of the thesis and in Appendix B.

Walshe, R.C. & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research, 100*, 38-46.
Appears in Chapter 2 within Part II of the thesis and in Appendix C.

The following conference proceedings paper does not appear as a chapter in the thesis. However, it contributed to the ideas developed within Chapter 1 (Part I) and is included in Appendix A:

Walshe, R.C. & Nuthmann, A. (2013). Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the* 35^{th} *Annual conference of the Cognitive Science Society* (pp. 1569-1574). Berlin, Germany: Cognitive Science Society.

ACKNOWLEDGMENTS

Above all, I would like to thank my wife Daniela Noll, you will always be my inspiration. I cherish your support, your kind words and your wonderfully creative mind.

I would also like to thank my parents, Ross and Catherine, and my brothers, Devon and Arran, who have collectively been my intellectual role models. Our frequent debates, rarely resolved, on every and all subject, are amongst my most formative childhood memories.

This thesis also owes a great debt to the many friends, new and old, who have gifted me their support and encouragement.

Lastly, I would like to give special thanks to my supervisor Antje Nuthmann who has been a considerate, patient and thought-provoking mentor.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Each day, a typical individual makes more than 100,000 fixations. These fixations serve many purposes, but they share a common goal; to support the biological capacity for vision. The ease and frequency with which these fixations are made hides the dazzling complexity of the underlying mechanisms that make such behaviours possible.

Although the visual field presents to the sensorium as a unified construct, this is far from reality for early visual processing. For example, when viewing a naturalistic scene, the image that is projected onto the retina is in constant flux. If considered individually, each fixation presents information to the visual system in a rather restrictive format. The foveated nature of the retina means that it is only the central region of the visual field that is represented in a high-resolution format. The foveated nature of human vision makes frequent fixations necessary for scene understanding.

The fact that the gist of a scene can be extracted in less than the amount of time required for a single fixation (Hillstrom, Scholey, Liversedge & Benson, 2012; Potter, 1976; Thorpe, Fize & Marlot, 1996) suggests that complex processing occurs even within these single fixations. But what mechanism does the visual system use to decide how long a fixation should be? This is an important question because if a fixation is consistently too short, the visual system will simply fail to acquire the features necessary to build up an understanding of what it is looking at. If fixations are too long, then this inefficiency may result in bottlenecks and a failure to detect behaviourally relevant information in a timely manner. It turns out that fixation durations appear to be remarkably well adapted to the type of task that they are engaged in. For example, fixation durations vary depending on whether an individual is searching for an object, reading, or viewing a naturalistic scene. These differences in the duration reflect the differing requirements placed on the visual system in these

tasks. However, eye-movements can also adapt to information that is presented within a single fixation. This level of adaptability allows eye-movements to be calibrated to directly match the properties of the stimulus that is currently being foveated.

To see how such a question might reflect eye-movements in the real-world, think about the act of searching for a set of keys in a messy room. Clearly this type of activity requires eye-movements. Because of the foveated visual system it is simply not possible to discriminate keys from non-keys when they are located in the periphery. Furthermore, the possible places at which a set of keys can reside goes well beyond the field of view. While scanning the scene, the visual system will encounter locations with many different spectral and object properties. Some regions will be low in luminance and contrast and some will be high in contrast with an average luminance. Some fixated regions will contain occluded or partial objects. Perhaps you will find a set of keys, but they will belong to someone else. All of these scenarios present different challenges for the visual system. Some fixations will require more time than others to complete visual processing. The theme that is developed in Part II of the thesis arises directly from such considerations. Using controlled experimental methods, an aspect of this real-world scenario is brought into the lab where measurements are made using an eye-tracking device. This is done by controlling the visual quality of a scene on a fixation-to-fixation basis. In all of the experiments in Part II, the aim is to design experiments in which the quality of a scene is enhanced, or degraded, without warning, on a single fixation. The impact that this change has on subsequent fixation duration is our primary measurement of interest. This case is not a simple one, but throughout this section we argue that there are some general eye-movement control principles that can be derived and which are later developed in Part III into a formal computational model.

Part I of the thesis provides an examination of a behaviour of fundamental concern to researchers interested in eye-movement control in high-level tasks. The well-founded assumption that eye-movements are sometimes modified, mid-preparation, has played a large role

in theorizing about eye-movements. That eye-movement modification happens is yet another example of the adaptive nature of the eye-movement control system. It is clearly advantageous for an eye-movement to have the ability to be changed midway through preparation. In a dynamic environment, objects tend to move position. What happens when the plan that has been hatched to move the eyes to an object suddenly becomes unnecessary because the object has shifted to a new position? Some 60 years of behavioural investigation using a technique known as the double-step approach has shown clearly that eye-movements can be updated mid-way through their preparation. However, there are caveats to eye-movement modification. Although a saccade can be modified, there is a limited window during which this is possible. This limitation may be summed up by saying that when an eye-movement is being prepared it passes a point-of-no-return; a point beyond which further modifications are possible. This point occurs roughly 80 ms prior to the onset of the eye-movement.

This exotic finding has had a large impact on theories of eye-movement control in high-level tasks. It can easily be seen why. In high-level tasks scenes frequently change. In certain instances, these changes will require an observer to update their viewing goals and strategies. However, if changing one's mind is only possible during certain epochs of eye-movement preparation, then this naturally has consequences for the behaviours that we observe. Therefore, understanding the measurements that we make during high-level tasks requires us to take into account this process of saccade modification. Part I of the thesis extends previous work in the study of saccade modification. Whereas previous studies used highly artificial stimuli (with good justification) this section of the thesis attacks the question of saccade modification from the perspective of a scene-viewing task. Previously, it was mentioned that the so-called point-of-no-return occurs roughly 80 ms prior to the onset of a saccade. However, does this change under certain circumstances? There is evidence from basic research on eye-movements that suggests that it does. Our question concerns whether this amount of time, which reflects the amount

of time required to modify an eye-movement program, depends on whether a person is viewing a naturalistic scene.

In Part III, like with Jeff Lebowski's proverbial rug, I attempt to tie the thesis together. Part I uses a basic psychophysical technique to explore eye-movement programming - specifically the timing of saccade modification processes. Part II takes a look at the question of eye-movements during scene perception from another vantage point. Here, the thesis looks at how features of a natural scene lead to changes in how long we look. The techniques in Part II tell us something about how eye-movements are extended to match processing requirements, but they don't reveal the underlying mechanisms as is done in Part I. Part III synthesizes these two approaches to studying eye-movements. Findings made in Part II of the thesis - along with findings made concurrently by other researchers (Glaholt, Rayner & Reingold, 2013; Henderson, Olejarczyk, Luke & Schmidt, 2014) - helped to shape a theoretical account of the mechanism that might explain these observed effects. In tandem with the low-level mechanisms revealed in Part I, these findings form the basis of the computational model that is formulated in Part III. In this part of the thesis simulations are presented to argue for one account of how eye-movements are controlled during naturalistic scene perception.

Part I

ON THE TIMING OF SACCADIC REACTIONS DURING NATURALISTIC SCENE PERCEPTION

Saccadic eye-movements are the primary vehicle by which human gaze is brought in alignment with vital visual information present in naturalistic scenes. Although numerous studies using the double-step paradigm have demonstrated that saccade preparation is subject to modification under certain conditions, this has yet to be studied directly within a naturalistic scene-viewing context. To reveal characteristic properties of saccade programming during naturalistic scene viewing, behaviour is contrasted across three conditions. In the Static condition of the main experiment, double-step targets were presented following a period of stable fixation on a central cross. In a Scene condition, targets were presented while participants actively explored a naturalistic scene. During a Noise condition, targets were presented during active exploration of a $1/f$ noise filtered scene. In Experiment 2, saccadic responses were measured in three static conditions (Uniform, Scene and Noise) where the backgrounds are the same as Experiment 1 but scene exploration is no longer permitted. It is found that the mechanisms underlying saccade modification generalize to both dynamic conditions. However, we show that a property of saccade programming known as the saccadic dead time (SDT), the interval prior to saccade onset during which a saccade may not be cancelled or modified, is lower in the Static task than it is in the Dynamic tasks. We also find a trend towards longer SDT in the Scene as compared to Noise conditions. We discuss the implication of these results for computational models of scene viewing, reading, and visual search tasks.

1

MECHANISMS OF SACCADIC DECISION MAKING WHILE ENCODING NATURALISTIC SCENES

1.1 INTRODUCTION

To access information rich regions of the visual field, the visual system engages in eye-movement behaviors known as fixations and saccades. The coordination of such movements involves a complex array of motor control mechanisms operating on distinct spatio-temporal scales. Fixations are defined as the state at which the eye remains in a relatively stable position on some aspect of the visual environment. In scene perception, it is known that the duration of fixations are dependent on such factors as the type of task that people are engaging in (Mills, Hollingworth, Van der Stigchel, Hoffman & Dodd, 2011; Nuthmann, Smith, Engbert & Henderson, 2010; Smith & Mital, 2013), the relevance of the fixated material to the task goals (Land & Hayhoe, 2001), as well as lower-level stimulus properties such as the luminance (Henderson, Nuthmann & Luke, 2013; Walshe & Nuthmann, 2014) and color (Ho-Phuoc, Guyader, Landragin & Guérin-Dugué, 2012) of the scene. Saccadic eye-movements are the primary mechanism by which the eye is brought into spatial alignment with scene content that is to be inspected in high-resolution foveal vision. Where observers fixate is influenced by both mid-level and higher-level stimulus properties (Nuthmann & Einhäuser, 2015). However, it is also known that the effect of image features on fixation selection in scenes may be overridden by task demands (Einhäuser, Rutishauser & Koch, 2008). Like fixation durations, properties of saccades are also known to vary depending on the task. For instance, participants generate larger saccade amplitudes when searching for an item than when memorizing a scene for later recall (Mills et al., 2011).

Theories of eye-movement control are primarily interested in elucidating the underlying, hidden mechanisms that generate behaviors such as saccades and fixations. The question of what event during stimulus processing results in the initiation of a saccade program to shift fixation away from the currently fixated location is a matter of considerable debate. This question has been most directly addressed in theories of eye-movement control while engaging in reading behavior where two competing views have been suggested (Reingold, Reichle, Glaholt & Sheridan, 2012, for review). The first view suggests that a saccade program is only triggered once the currently viewed stimulus has been processed to a sufficient degree (Reichle, Pollatsek, Fisher & Rayner, 1998; Reichle, Pollatsek & Rayner, 2012). Variants of this view are referred to as *cognitive trigger theories*, as saccadic eye-movements are only generated when cognition-related processing has been achieved to a sufficient depth. In contrast to the triggering mechanisms just described are those that suggest that the variability in the termination of a fixation is a result of difficulties in lexical processing that interfere with the saccade initiation processes. In such models, the decision to initiate a saccade is achieved by an autonomous random timer, and the duration of this timing process may be modulated by the difficulties encountered during stimulus processing (Engbert, Nuthmann, Richter & Kliegl, 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Therefore, moment-to-moment difficulties in processing result in increased random timing intervals, and consequently, longer fixation durations. Such principles of saccade generation have been adapted to explain fixation duration distributions within complex scene-viewing tasks. The CRISP model (Nuthmann et al., 2010) incorporates a random-walk timer that generates signals to begin the programming of a saccade. Importantly, the CRISP model also allows such saccade programs to be cancelled in the case that programming has not proceeded to a sufficiently advanced stage. The CRISP model has been demonstrated to provide good fits to empirical data under a number of experimental conditions (Nuthmann et al., 2010; Nuthmann & Henderson, 2012).

A powerful tool that has been used to reveal the empirical properties of saccade programming timelines is known as the double-step paradigm (Westheimer, 1954; Becker & Jürgens, 1979). Classic variations of the double-step paradigm involve presenting participants with two targets along a horizontal axis with a varying interstimulus interval separating the two targets. For instance, in a seminal study Becker & Jürgens (1979) had a condition in which a first target was presented at 15° to the left or right of fixation with a second target presented at 30° in the same direction at delays of 50, 100, 150 and 200 ms. The participants' task was to fixate the target as quickly as possible which thereby placed a saccade program that had been initiated to target the first location in competition with a saccade program targeting the second location. The critical behavioral measure in this study was the amplitude of the saccade that was observed in response to the presentation of the double-step stimulus. Responses observed to fall spatially close to the location of the initial target step permitted the inference that information regarding the updated target location was not incorporated into the programming of the saccadic response. Likewise, the degree to which saccade amplitudes deviated from the initial location and tended to land close to the second target provided a measurement of the degree to which the second target had influenced the resulting saccade. The authors observed an interesting temporal dependency between the amplitude of the saccadic response and presentation of the second target. If the response saccade occurred in close temporal proximity to the appearance of the second target, then the response saccade tended to fall close to the initial target position. As the temporal interval between the presentation of the second target and the response increased, responses gradually tended towards the final target position. These results revealed that saccades may be reprogrammed when updated position information is available, and that the mechanisms underlying this behavior are dependent on the temporal relationship between the updated target position and the onset of the response saccade. This finding has since been replicated by many authors (Findlay & Harris, 1984; Aslin & Shea, 1987; Ludwig, Mildinhall & Gilchrist, 2007;

Camalier, Gotler, Murthy, Thompson, Logan, Palmeri & Schall, 2007; Ray, Schall & Murthy, 2004).

A question that follows directly from such results is at what point in time prior to the onset of a saccade does updated target position information cease to have an influence on the impending saccade? Such a point of no return is often referred to in the literature as saccadic dead time (SDT). Findlay & Harris (1984) analyzed double-step data from a replication of Becker & Jürgens (1979) and found that saccades begin to incorporate targets that are displaced 80 ms prior to saccade onset. More recently, Ludwig et al. (2007) used a double-step task to investigate whether estimated values for SDT show evidence for sensitivity to task effects. In their study, the angle of displacement between the first and second double-step target was manipulated. They found that SDT increased as the angle between the initial and final target locations increased. These results provide empirical support for the assumption that SDT values are sensitive to characteristics of the task environment.

The concept of a point of no return has been suggested as an explanation for a number of empirical effects observed within the scene-viewing literature. In a procedure known as the stimulus onset delay paradigm (SOD) (Shioiri, 1993; Henderson & Pierce, 2008) participants viewed naturalistic scene stimuli while engaged in an encoding task. During a critical fixation, the scene was removed and replaced by a mask, thereby removing scene content from further processing. Within the same fixation, the scene was then restored to view at varying delays. Results from the SOD paradigm showed a bimodal distribution of fixation durations such that one population was independent of the length of the delay whereas the other population increased in proportion to the length of the delay. In the context of the CRISP model, Nuthmann et al. (2010) suggested that such bimodality arises due to two factors. First, when the scene disappears from view the rate at which saccade programs are generated slows down to reflect the lack of incoming visual information. Secondly, if a saccade program is currently within the labile stage of saccade programming (i.e., has not passed the point of no return) then the current

saccade program is subject to stochastic cancellation. Together, these processes combine to yield delays in the programming of saccades and therefore longer fixation durations. Wu, Gilani, van Boxtel, Amihai, Chua & Yen (2013) also demonstrated an influence of a point of no return on saccade programming in a study in which a scene was swapped with a novel scene while participants were engaged in a viewing task. In the saccade that immediately followed the change, it was found that participants systematically programmed saccades to target the center of the screen. However, such a center scene bias occurred primarily for saccades that were initiated at relatively long durations after the change. For saccades that were initiated shortly after the display change, there was no evidence for such a bias. Similar to the SOD paradigm, such a result is well accounted for by the fact that in those saccades that did not target the center, the scene change occurred when saccade programming had already passed the point of no return and could therefore no longer influence saccade programming.

The principles derived from such investigations with double-step stimuli have provided the basis for the implementation of eye-movement control models in scene viewing as well as in reading. In the E-Z Reader model (Reichle et al., 1998), the authors introduced the concepts of a labile and a nonlabile stage of saccade programming. This distinction implies a multi-stage saccade programming architecture whereby saccade programming that is within the labile stage is subject to cancellation. However, once programming has progressed beyond the labile stage into the nonlabile stage, cancellation is no longer possible. Such multi-stage saccade programming assumptions have since been incorporated into a number of models that explain oculomotor control under a variety of task conditions such as scene viewing (Nuthmann et al., 2010), reading (Reichle et al., 1998; Pollatsek, Reichle & Rayner, 2006; Engbert et al., 2005), and visual search (Trukenbrod & Engbert, 2014).

One notable difference between these models are the assumptions that are made regarding the duration of the nonlabile stage of saccade programming. In both the CRISP model (Nuthmann et al., 2010)

and the E-Z Reader model (Reichle et al., 1998) it is assumed that the nonlabile stage adopts a fixed duration and is not dependent on stimulus characteristics. On the other hand, as of SWIFT-II (Engbert et al., 2005), the model makes the explicit assumption that the duration of the nonlabile stage may vary depending on the amplitude of the planned saccade. Therefore, the SWIFT model explicitly incorporates systematic variability in SDT while the CRISP model and E-Z Reader do not. It is important to note that all models just introduced do incorporate *random* (i.e., unsystematic) variability in the duration of the nonlabile stage as the duration of a given nonlabile stage is drawn from a statistical distribution. While Ludwig et al. (2007) provides some empirical support for such stimulus dependent differences in the duration of the nonlabile stage, there is currently no empirical research demonstrating task-specific differences within reading or scene-viewing contexts. Furthermore, upon inspection of the model parameters that are used to describe the duration of the nonlabile stage, it becomes evident that there is very little consistency between (and within) models regarding the duration of this stage. We return to this issue in the General Discussion.

The aim of the current study is to investigate saccade programming during naturalistic scene perception by embedding a double-step task within scene-viewing contexts. Two primary questions are addressed with these experiments. The first addresses an empirical gap in the scene-viewing literature by testing whether saccade cancellation operates in an analogous manner within scene viewing as it does within classic double-step investigations. The assumption that findings from low-level tasks generalize to high-level task contexts has often been made by models of oculomotor control (Reichle et al., 1998; Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). However, it is an open question that has remained largely unaddressed (but see Walshe & Nuthmann, 2013). Our second question asks whether SDT values vary depending on experimental context. By deriving such an estimate we also explore how such a value may provide an indirect measure of the duration of the nonlabile stage by taking into account the delays in transmission of informa-

tion between the retina and cortical regions responsible for saccadic decisions.

In Experiment 1 we address these questions by comparing double-step performance across three conditions. The first condition (Static) replicates a classic version of the double-step procedure (Becker & Jürgens, 1979). In a second condition (Scene), a scene-viewing double-step task is conducted by presenting the double-step targets during active exploration of scene content. The third condition (Noise) replicates the experimental design of the scene-viewing double-step task, but instead replaces the scene with a phase noise transformed background stimulus. Such a transformation is achieved by applying noise to the phase spectrum, but leaving the amplitude spectrum intact (Einhäuser, Rutishauser, Frady, Nadler, König & Koch, 2006). Applying a transformation in this manner removes object and other higher order scene statistics yet retains the $1/f$ characteristics of the amplitude spectra that is typical of naturalistic scenes (Einhäuser et al., 2006). In a followup experiment we test how SDT is influenced by scene background independently of dynamic movement context. In Experiment 1, observed differences in SDT between the Static condition and the dynamic movement conditions (Scene, Noise) could be attributed to either the background content or differences in the movement context. Experiment 2 provides a test of these two hypotheses by comparing the three backgrounds used in Experiment 1 in static movement contexts only.

To make comparative inferences about saccade programming across conditions, we utilize a non-linear mixed-effects (*nlme*) regression framework that improves upon previous methods of analyzing double-step performance. Population-level parameters provided a method to compare performance across conditions, while individual parameter estimates were extracted from the fitted model and were used to provide a by-participant measure of SDT in the three conditions.

1.2 EXPERIMENT 1

1.2.1 *Methods*

Participants

Two males and 13 females (mean age: 23 years) recruited from the University of Edinburgh student population participated in the study. Participants completed all experimental conditions in one session, which lasted approximately 1.5 hours. Each participant was paid £7 per hour of participation in compensation for their time. The study conformed to the tenets of the Declaration of Helsinki, and written consent was supplied by the participants prior to the experiment.

Apparatus

Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz and the monitor screen was at a distance of 67 cm from the participant. During stimulus presentation, participants' eye-movements were recorded using an SR Research EyeLink 1000 Desktop mount system. It was equipped with the 2000 Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz for each eye with an average spatial accuracy of 0.25°-0.5° of visual angle. Viewing was binocular, and both eyes were tracked. Only the position of the right eye was used in the analysis. A chin rest was used to achieve stability of a participants' head position relative to the screen. The experiment was implemented in MATLAB 2009b using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray & Broussard, 2007), which incorporates the EyeLink Toolbox extensions (Cornelissen, Peters & Palmer, 2002). The software allowed precise control over the timing of display changes. To detect fixations online, we implemented a 9-sample online velocity estimation algorithm in MATLAB that aimed to mimic Data Viewer's offline velocity estimation procedure (SR Research Ltd., 2006). Fixations were detected offline utilizing SR Re-

search Data Viewer to parse the gaze samples into sequences of fixations and saccades.

Stimuli

In all conditions the targets consisted of isoluminant 1.5° square boxes presented in the color pink, (CIE_{Lab} L = 65.48, a = 61.84, b = -26.03). In the Static task, the background was uniformly black. When participants were required to fixate on a central cross, it was presented in red (CIE_{Lab} L = 53.23, a = 80.42, b = 66.96). In the Scene task, participants viewed images of 200 naturalistic scenes, in addition to 4 practice scenes. Each scene had a resolution of 800×600 pixels and was presented in full color. Scenes were collected from online databases such as flickr and google images. They were selected to include a variety of categories such as indoor and outdoor as well as urban and nature scenes. Each scene was viewed by the participant only once over the duration of the experiment. At a viewing distance of 67 cm, the scenes subtended $33° \times 25°$. The stimuli in the Noise task consisted of 200 background images that were constructed by applying a phase noise filtering procedure to the images used in the Scene task. The application of the phase noise filtering process allowed removal of higher order scene statistics such as those used to determine edges and contours while leaving the amplitude spectrum unmodified. Phase noise stimuli were created by transforming the original scene images into Fourier space where additive noise drawn from a uniform distribution was added to the phase spectrum (Einhäuser et al., 2006). An inverse Fourier transformation was then applied to the images to convert them back to image space.

Procedure

Three double-step tasks were conducted to compare characteristics of saccade programming in static and dynamic gaze conditions. The order of the conditions was counterbalanced across participants. In each of the three tasks, participants completed 200 trials that were classified as either double-step (40%) or single-step (60%) trials. Single-step trials were included to ensure that participants could not make

accurate anticipatory saccades to the final resting location of the double-step target. On single-step trials, the center of the 1.5° pink box was presented at an eccentricity of 7° from the point at which the participant was currently fixating. The order of single-step and double-step trials was randomized.

On double-step trials, two targets were presented at distances of 7° and 14° of visual angle from the fixated location. The first target is referred to as the *initial target* and the second target as the *final target*. Both targets were always presented in the same direction relative to the fixation location. Therefore, since the initial target was presented at 7° from the fixation location, the final target was always presented at a further 7° in the same direction. The color and luminance of the stimuli in the double-step condition were identical to that of the target in the single-step condition. The timing of the initial target step was identical to that of the timing of the single-step stimuli described above. The initial target step was presented for varying amounts of time prior to the onset of the final target step. The interval defining the amount of time elapsed between the onset of the first target step and the onset of the second target step is referred to as the *target step delay* (TSD). The final target was presented simultaneously with the disappearance of the first target. Therefore, the subjective impression of this procedure is that the first target step *jumps* to the second target location.

The TSD for a trial was defined in an adaptive manner such that TSD varied depending on the amplitude of the response on the previous double-step trial (Camalier et al., 2007). A *compensated saccade* refers to a saccade that was programmed to go directly to the final target location, whereas a *non-compensated saccade* is one in which the saccade was programmed to the initial target location. From previous investigations, it is known that short TSDs tend to result in final target response saccades whereas longer TSDs tend to result in initial target responses (Becker & Jürgens, 1979; Camalier et al., 2007). Following a double-step trial in which a compensated saccade was produced, TSD was increased, and TSD was decreased following non-compensated saccades. The adaptive increment was defined as

Figure 1: Experiment Design: In **A** an example stimuli from the 1/f Noise condition is presented. The procedure for the double-step condition in the dynamic tasks is illustrated in **B**. The experiment begins with the participant fixating on a central cross. The participant then views a scene (or 1/f) stimuli for later recall. On the 10^{th} fixation the first target is presented immediately upon detection of fixation onset, followed by the second target at intervals of 50, 100, 150, or 200 ms. An adaptive staircase method was used to define the interval on a specific trial. The dashed blue circle represents an example fixation location before target presentation, and the red circle represents an example fixation location after a target response is generated. A *compensated saccade* occurs when a saccadic response is generated directly to the final position of the target (top left of panel **B**). A *non-compensated saccade* occurs when a response is erroneously generated to the first target position (top right of panel **B**). In panel **C** the single-step condition is illustrated, in which only a single target is presented. Panel **D** illustrates the procedure for a double-step trial in the Static condition. Participants fixate on a red cross at the centre of the screen. Targets then appear at intervals of 2000-3000 ms. Trials can either be single-step or double-step trials, and the timing is the same as in panels **B** and **C**.

50 ms; a lower bound on TSD was set at 50 ms. The purpose of this adaptive procedure was to balance the number of compensated and non-compensated response saccades. Compensated saccades were detected online and were identified when a saccade was made within 2° of the final target. The presentation of the stimulus was synced with the vertical retrace of the monitor. Full presentation of the stimulus was therefore delayed from the defined TSD by up to 7.14 ms.

The trial was terminated 1000 ms following the first saccade made in response to the presented targets.

Static Task

In the Static condition, participants initiated a trial by fixating on a cross presented at the center of the screen. If the eye-tracker could not reliably detect fixation on the cross, a recalibration routine was initiated. Following a randomly timed delay of 2000-3000 ms, targets were placed on the same horizontal axis as the fixation cross and were presented either to the left or to the right. On double-step trials, the final target was always presented on the same side of the cross as the initial target. Targets were presented to the left or right with equal frequency and the side was randomly selected for a particular trial. Participants were instructed to fixate the box as quickly and accurately as possible. The design in the Static condition is visualized in Figure 1.

Dynamic Tasks

In the dynamic tasks (Scene, Noise) participants initiated a trial by fixating on a central cross. Recalibration of the eye-tracker was initiated if a reliable fixation on the cross could not be detected. Once the cross had been fixated, the image was revealed and the participant was free to make unconstrained fixations on the screen. In both conditions, the task was to encode the presented image for later recall. Participants were instructed that the recall phase would commence once all the stimuli had been presented. Once the participant had made 10 saccades on the current trial, a critical fixation was identified and the targets were presented. The targets were presented immediately upon detection of the onset of a critical fixation. Fixations were detected online using a custom 9-point velocity estimation algorithm (see Apparatus). A primary difference between the static and dynamic tasks is the trajectory at which the targets were presented. In contrast to the Static task, in the dynamic tasks the targets could be presented along any axis and the location of the targets was determined by the position of the current and preceding fixation lo-

cations. The targets were presented along an axis that was derived by intersecting the coordinates of the current fixation with the previous fixation. The initial target was therefore placed at a distance of 7° from the currently fixated location in the same direction as the most recent saccade (see Figure 1 for details). The decision to place the targets along such a trajectory was done to control for systematic viewing biases that are known to exist during the exploration of naturalistic scenes. Firstly, it is known that during scene-viewing tasks saccades are most frequently programmed in the horizontal direction (Tatler & Vincent, 2008; Foulsham, Kingstone & Underwood, 2008; Nuthmann & Henderson, 2010). We confirmed that such a systematic bias exists in our dynamic conditions by visual inspection of radial histograms showing saccade angles relative to the horizon (see Figure 2). To account for such a horizontal bias, we implemented a static double-step task in which targets were presented exclusively on the horizon. This was done to maximize the match between target trajectories in the static and dynamic conditions. Secondly, we wanted to control for the effect of *saccadic momentum*. Saccadic momentum refers to the tendency for fixation durations to increase as the angle between the preceding and subsequent saccade direction increases (Wilming, Harst, Schmidt & König, 2013). For example, saccades that are programmed perpendicular to the direction of the preceding saccade result in longer fixation durations. Therefore, by placing both targets along the same trajectory, any influence on the timing of the saccade due to saccadic momentum is minimized. The design in the dynamic tasks is visualized in Figure 1.

1.3 RESULTS

Prior to analysis, the data were pre-processed to exclude cases that did not conform to sufficient data quality standards. Each participant viewed 200 trials in each condition, 40% (80) double-step trials and 60% (120) single-step trials. In double-step trials, we excluded any trial in which the response saccade was initiated prior to the time at which both targets had been presented. This left an average of

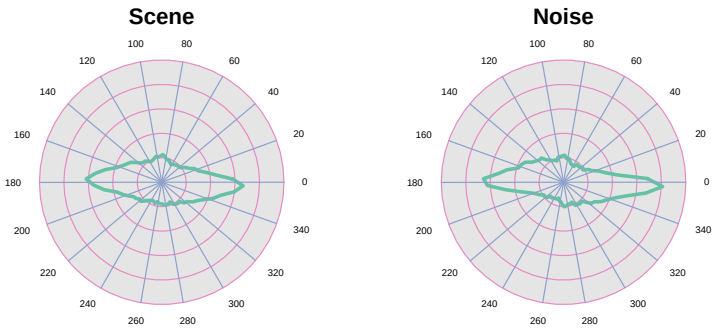

Figure 2: Distribution of saccade directions. In both the Scene and Noise conditions there is a clear preference to program saccades along the horizontal axis. Angles of 0° and 180° indicate saccades programmed along the horizon. Densities were calculated from a bin size of 7.2°.

61 double-step trials in the Static condition, 74 in the Scene condition, and 72 in the Noise condition. Saccades that were clearly too short (< 2°) to be considered as responses to either of the targets were excluded. Furthermore, if a blink occurred immediately before or immediately after the fixation in which the targets were presented then this trial was excluded. In the Static task, if the targets were presented when the fixation deviated by more than 2° from the fixation cross then this trial was excluded. An additional criterion was defined such that if a saccade was programmed more than 45° away from the direction that the targets were presented in, then these saccades were considered to not be programmed in response to detection of the targets and were excluded from the analysis. After all exclusion criteria had been applied there remained an average of 56 double-step trials in the Static condition, 67 trials in the Scene condition, and 68 trials in the Noise condition.

1.3.1 *Modeling the Amplitude Transition Function*

The amplitude transition function (ATF) relates the resulting saccadic response amplitude to a quantity referred to as Delay (D). D measures the time elapsed between the onset of the second target step and the onset of the response saccade (Becker & Jürgens, 1979). The ampli-

tude of the saccade provides information about whether a saccade was successfully reprogrammed or not. In the present experiment, amplitudes of approximately 7° indicate saccades programmed to the initial target while amplitudes greater than 7° can provide evidence for the saccade being executed towards the final target location.

Values of D provide a measurement of how much time has elapsed between the onset of the second target and the onset of the saccadic response. Therefore, D measures the amount of time available to reprogram a saccade to the new target location. If the value of D is high, this implies that the second target was available for a relatively long period prior to the onset of the saccade. When the value of D is low, the saccade was executed very shortly after the presentation of the second target. By combining D with the amplitude of the response, thereby constructing an ATF, it is possible to ask the following question: What is the minimal amount of time prior to the onset of the saccade that the second target must be presented to have an influence on the resulting saccade?

Previous research has demonstrated that the ATF in double-step tasks may be well described by a curve that closely resembles the logistic function (Becker & Jürgens, 1979; Ludwig et al., 2007). This was confirmed by graphical analysis of the data in the present experiments as well as ATFs previously reported in the literature. The following four-parameter logistic function was used to model response amplitude as a function of D in the three experimental conditions:

$$f(x) = \alpha + \frac{\beta - \alpha}{1 + e^{\gamma(\delta - x)}}, \tag{1}$$

where α represents the lower bound for the logistic function, β represents the upper bound, γ is a scaling parameter, and δ defines the inflection point.

As a novel approach, ATFs were estimated with a non-linear mixed-effects regression framework using the *nlme* program of the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar & R Core Team, 2014) implemented in the R software for statistical computing (R Development Core Team, 2012). Using this approach we model the variability in

the estimation of parameters contributed by both individual participants and experimental condition. Specifically, a non-linear mixed-effects model provides a method of simultaneously estimating the fixed (population) level parameters and the random (individual-level) parameters (Pinheiro & Bates, 2000). A benefit of estimates derived from such a framework is that participant-level parameter estimates are weighted by their corresponding population parameters, providing a measure of protection against overfitting the ATFs to individual-level data. By-participant random effects (intercept and slope) were included for all four parameters of the logistic function, thereby allowing all parameters of the model to vary in a by-participant manner. The random effect covariance matrix was assumed to follow a block diagonal structure. That is, correlations between random intercepts and slopes were only permitted when grouped within the logistic parameters α, β, γ, δ and were assumed to be 0 otherwise. For example, correlations between the upper bound intercept and slopes were estimated, but correlations between the upper and lower bound random effects were assumed to be 0. Parameters were estimated by maximizing the log-likelihood of the model given the observed responses. The parameters of the fitted model are presented in Table 1.

1.3.2 *Saccade Modification*

The ATF is a direct way to measure the behavioral correlates of mechanisms underlying saccade reprogramming as it measures the amplitude of saccades that are initiated under conditions in which updated target position information is available. Furthermore, the ATF plots the response amplitude as a function of the duration that has elapsed since the presentation of the second target D. Therefore, the ATF allows the relationship between the time spent processing the second target and the amplitude of the response to be revealed. The ATF allows specific predictions about saccade cancellation processes to be tested. In the case that saccades can never be modified, the ATF predicts a flat function with an intercept close to the location at which the first target is presented. In the present experiment, the predicted

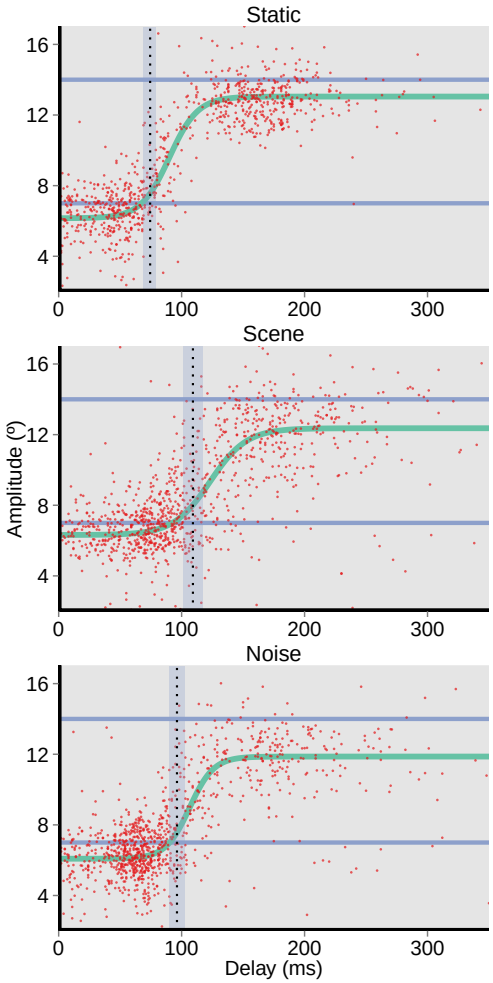

Figure 3: Amplitude Transition Functions for the three experimental conditions in Experiment 1. Points represent the amplitude of saccades initiated in response to the presentation of double-step targets. On the x-axis the Delay (D) represents the amount of time elapsed between the presentation of the second target and the onset of the saccade. The y-axis represents the amplitude of the resulting saccade. The green lines represent the best fitting fixed effects curves estimated with a non-linear mixed-effects regression. The horizontal solid blue lines represent the physical distance of the targets in relation to fixation location (7° and 14°). The vertical dotted black lines represent the mean SDT in each condition derived from the by-participant SDT estimates. The light blue bands surrounding the means are bootstrapped confidence intervals.

intercept would be 7°. In the case in which modification is always possible, a flat function would also be predicted, but in this case the intercept would be predicted to be near the location at which the second target is presented; 14° in the present experiment. An intermediate hypothesis between these two extremes is that modification of a saccade program is possible and becomes increasingly likely at

Table 1: ATF Regression Model (Experiment 1)

Fixed Effects	Parameter	Estimate	Std. Error	t-value
Lower Bound	Static (Intercept)	6.15	0.19	31.56
(α)	Noise	-0.07	0.18	-0.39
	Scene	0.164	0.173	0.946
Upper Bound	Static (Intercept)	13.05	0.15	85.62
(β)	Noise	-1.18	0.21	-5.50
	Scene	-0.68	0.24	-2.81
Inflection	Static (Intercept)	90.05	2.30	39.10
(δ)	Noise	16.49	2.31	7.14
	Scene	33.10	2.10	11.04
Scale	Static (Intercept)	0.09	0.01	8.39
(γ)	Noise	0.07	0.01	1.94
	Scene	-0.02	0.02	-1.26

Random Effects	Parameter	σ
Lower Bound	Static	0.59
(α)	Noise	0.42
	Scene	0.20
Upper Bound	Static	0.47
(β)	Noise	0.58
	Scene	0.70
Inflection	Static	6.77
(δ)	Noise	4.15
	Scene	6.80
Scale	Static	0.03
(γ)	Noise	0.01
	Scene	0.05
Error Term	ε	1.64

Estimated parameters for the four-parameter logistic regression model in Experiment 1. Means, standard errors, and t-values of fixed effects; standard deviations of the random effects.

greater temporal separation between the onset of the second target and the onset of the saccadic response. In this case, a monotonically

Table 2: Estimated Saccadic Dead Time (ms) in Experiment 1

Participant	Static	Scene	Noise
1	79	115	94
2	57	127	95
3	67	103	83
4	75	89	89
5	73	105	96
6	83	111	100
7	85	115	106
8	79	112	100
9	69	115	92
10	83	118	114
11	86	134	115
12	70	90	85
13	65	90	85
14	84	122	104
15	65	95	85
Mean	74	109	96

increasing ATF would be predicted with a lower asymptote located close to the first target location that gradually increases and asymptotes close to the final target location. Inspection of Figure 3 reveals that the shape of the ATF in all three experimental conditions conforms to the intermediate hypothesis. Saccade programs can be modified and as greater values of D are observed, a saccade targeting the final location becomes increasingly likely.

Furthermore, the regression modeling reveals that, for double-step trials, saccades significantly undershoot the targets and that the degree of undershoot depends on experimental condition. The Static condition was selected as the intercept for each of the logistic parameters that were estimated. Effects were determined to be significantly different from 0 when $|t| > 1.96$ was observed. For the lower bound, the estimated amplitudes for all conditions were less than the distance at which the targets were placed (7°). Task did not not significantly influence the lower bound. That is, the estimated change in lower bound from the Static condition (intercept) was not significant for either the Noise or Scene condition. Significant task effects were

observed for the estimated upper bound. The upper bound estimate for the Static condition (intercept) was $\hat{\beta} = 13.05°, t = 85.62$, which indicates an undershoot to the second target. The decrease in upper bound was greatest in the Noise condition, $\hat{\beta} = -1.18°, t = -5.50$, followed by the Scene condition, $\hat{\beta} = -0.68°, t = -2.81$. See Table 1 for details.

1.3.3 *Saccadic Dead Time*

Saccadic dead time may be defined as the last point in time at which novel stimulus information may be incorporated by the system responsible for preparing a saccadic response. Thus, SDT may also be described as the point of no return in the preparation of a saccade. Once the point of no return in saccade programming has been reached, that saccade may no longer be modified or cancelled.

To estimate the SDT from the fitted ATFs, we determined the largest value of D for which the amplitude of the response saccade showed no evidence of incorporating the second target position into the response. To define saccade amplitudes in the double-step condition that are inconsistent with responses programmed to the first target location, we used the distribution of response amplitudes observed in the single-step condition. In the single-step condition, saccades target the first location and are characterized by a distribution of amplitudes that cluster near the first target location. For each participant and each condition we measured the amplitude that corresponds to the 95^{th} percentile of responses. Such a value provides a cutoff for amplitudes that are rarely associated with responses to the initial target location. This cutoff was used to define an amplitude threshold in the double-step condition such that responses with an amplitude beyond this point were considered to be influenced by the second target step[1].

As the ATF was estimated with a non-linear mixed-effects regression with by-participant random intercepts and slopes of experimen-

1 We thank Casimir Ludwig for suggesting this method. We also note that calculating SDT in such a manner resulted in estimates that were increased in comparison to an alternative method (see Ludwig et al., 2007).

tal condition, it was possible to provide an independent estimate of SDT for each participant in each condition. Specifically, estimates of SDT were computed for individual participants by inspecting the responses predicted by the individual-level data from the mixed-effects regression. Individual-level SDT estimates, along with mean SDT in the three conditions, are provided in Table 2. To statistically validate the differences in observed means, we utilized a bootstrap procedure to estimate the distribution of mean SDT in each of the three conditions (Efron & Tibshirani, 1993). In summary, the SDT observations recorded in each condition were resampled with replacement 10,000 times, and for each iteration a sample mean was recorded. Through this method, we constructed a distribution of sample means based on bootstrapped data. These distributions were then used to construct 95% confidence intervals (CI) around the observed means. We defined a significant difference between means in any two conditions as occurring when the bootstrapped CIs did not overlap. To control for family-wise error rates Bonferroni correction was applied to the CIs. The mean SDTs and CIs were derived from the fitted model. The mean SDT was shortest in the Static condition ($M = 74$ ms, $CI_{low} = 69.03$, $CI_{high} = 79.65$), followed by the Noise condition ($M = 96$ ms, $CI_{low} = 90.49$, $CI_{high} = 102.69$) and the Scene condition ($M = 109$ ms, $CI_{low} = 100.94$, $CI_{high} = 117.27$). Therefore, according to the bootstrapped hypothesis testing procedure, statistically reliable differences were observed between SDT in the Static vs Scene and Static vs Noise conditions. Furthermore, a trend was observed in the Scene vs Noise conditions such that SDT tended to be longer in the Scene than in the Noise condition. Inspection of the individual measures of SDT (see Figure 4) shows that the SDT for each participant was numerically larger in the Scene than in the Noise condition.

1.3.4 *Additional analyses*

Additional analyses explored whether the structural differences in the stimulus content in the two dynamic tasks were associated with

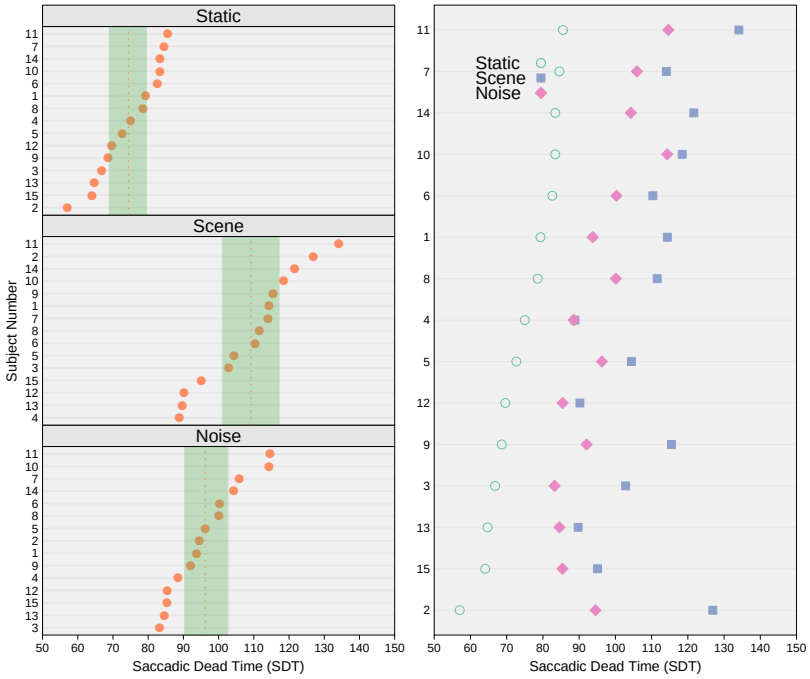

Figure 4: Saccadic Dead Time (SDT) estimates (ms) for the three experimental conditions in Experiment 1. The left plot shows by-participant estimates of SDT. For a given condition, from bottom to top, values are ordered from lowest to highest. The vertical dotted lines represent the mean value of SDT in each condition. The green band surrounding the mean is the bootstrapped confidence intervals of the mean estimate. The right plot shows participants ordered by the magnitude of their estimated SDT value in the Static condition.

differences in global eye-movement parameters. To compare mean saccade amplitudes and fixation durations in the Noise and Scene conditions, only those saccades that were not generated in response to a target step were considered. For saccade amplitudes, no statistically significant difference was found ($p > 0.05$). Saccade amplitudes to single-step targets were also analyzed. They were shortest in the Noise condition ($M = 6.20°$), followed by the Static condition ($M = 6.44°$) and the Scene condition ($M = 6.51°$). Pairwise t-tests with a family-wise error rate of 0.01 were used to statistically validate the mean differences. The difference between Scene and Noise was statistically significant, $t(14) = 3.49$, $p = 0.004$, as was the difference between Static and Noise, $t(14) = 2.80$, $p = 0.01$. The difference between Static and Scene did not reach statistical significance.

Mean fixation durations were longer in the Noise Condition ($M =$ 324 ms) compared with the Scene condition ($M = 276$ ms), and this

difference was significant, $t(14) = 4.28$, $p = 0.0008$. This pattern of results replicates a previous study that compared $1/f$ filtered scenes with naturalistic scenes (Kaspar & König, 2011).

Additionally, we analyzed saccade reaction times to single-step stimuli. Saccadic responses were fastest in the Noise condition ($M = 161.82$ ms), followed by the Scene ($M = 176.59$ ms) and Static ($M = 194.83$ ms) conditions. Saccadic response times were significantly longer in the Scene than in the Noise condition, $t(14) = 2.51$, $p = 0.03$. The difference between the Static and the Noise condition was also significant, $t(14) = 4.19$, $p = 0.001$. The difference between the Static and the Scene condition did not reach statistical significance.

1.4 EXPERIMENT 2

A second experiment was conducted to compare SDT under conditions in which the structure of the background is varied but in which participants do not explore the scene prior to target onset. In this experiment, participants conducted three versions of the Static task from Experiment 1. In the first condition, the task is conducted on a uniformly black background. We call this condition the Uniform condition and note that it is identical to the Static condition in Experiment 1. In the second condition (Scene condition) the same task is conducted, but the background is replaced by one of the naturalistic scenes used in Experiment 1. In a final condition, the task is conducted with $1/f$ stimuli presented as the background.

As movement is controlled in this study, any observed difference in SDT can be more directly interpreted to result from differences in the structure of the background. Specifically, we predict that if increases in SDT observed in Experiment 1 are due to the additional structure of the backgrounds in the Scene and Noise conditions, then the structured background conditions in Experiment 2 should reveal elevated SDT relative to the uniform background.

The experiment was conducted on an additional 7 male (including one author, RCW) and 5 female participants with an average age of 24 years who did not participate in Experiment 1. The three conditions

in Experiment 2 (Uniform, Scene, Noise) paralleled in nearly all details the Static condition from Experiment 1. The primary difference was that in two of the conditions, the backgrounds were replaced by scenes (Scene condition) or phase noise images (Noise condition) that were used in Experiment 1. On each trial, the specific scene or noise image was randomly selected from the set of stimuli used in Experiment 1. Participants were explicitly told that the scene or noise background content was not relevant to completion of their task.

1.4.1 *Results*

The approach to the analysis of the results in Experiment 2 was conducted in an analogous way to the Static condition in Experiment 1. The non-linear mixed-effects model in Experiment 2 included the same random effects structure as the model in Experiment 1. That is, by-participant random intercepts and slopes were included for α, β, γ and δ. The model estimates are summarized in Table 3.

The estimated SDT in the Uniform condition was $M_{Uniform} = 71.39$ ms, $CI_{low} = 65.00$, $CI_{high} = 78.26$. In the Scene condition, SDT was estimated at $M_{Scene} = 80.70$ ms, $CI_{low} = 75.50$, $CI_{high} = 86.11$ and was estimated at $M_{Noise} = 70.83$ ms, $CI_{low} = 65.93$, $CI_{high} = 76.55$ in the Noise condition. As was introduced in the analysis of Experiment 1, lack of overlap in Bonferroni corrected confidence intervals was used as a criterion for rejecting the null hypothesis. Accordingly, there was no difference in SDT between the Uniform and Noise conditions, as is evidenced by the strongly overlapping CIs (Figure 6). The CI for the Scene condition slightly overlapped with the other two CIs. However, inspection of the pattern of results at the level of individual SDT responses revealed a strong trend for longer SDT in the Scene condition when compared to the Uniform and Phase conditions. For 10 out of the 12 participants, the longest SDT response was observed in the Scene condition (Figure 6, Table 4).

We also analyzed saccade latencies on trials in which only the initial target was presented. Latency was lowest in the Uniform condition ($M = 203.29$ ms), followed by the Noise condition ($M = 207.92$

Table 3: ATF Regression Model (Experiment 2)

Fixed Effects	Parameter	Estimate	Std. Error	t-value
Lower Bound	Uniform (Intercept)	6.55	0.15	44.15
(α)	Noise	0.37	0.13	2.96
	Scene	0.35	0.12	2.92
Upper Bound	Uniform (Intercept)	13.06	0.16	79.48
(β)	Noise	-0.08	0.14	-0.57
	Scene	-0.10	0.13	-0.74
Inflection	Uniform (Intercept)	86.62	3.79	22.88
(δ)	Noise	2.71	2.90	0.94
	Scene	9.96	2.31	4.32
Scale	Uniform (Intercept)	0.10	0.01	7.83
(γ)	Noise	-0.01	0.01	-1.04
	Scene	-0.004	0.02	-0.22

Random Effects	Parameter	σ
Lower Bound	Uniform	0.42
(α)	Noise	0.10
	Scene	0.02
Upper Bound	Uniform	0.52
(β)	Noise	0.33
	Scene	0.34
Inflection	Uniform	12.63
(δ)	Noise	8.63
	Scene	6.34
Scale	Uniform	0.04
(γ)	Noise	0.02
	Scene	0.04
Error Term	ε	1.11

Estimated parameters for the four-parameter logistic regression model in Experiment 2. Means, standard errors, and t-values of fixed effects; standard deviations of the random effects.

ms) and Scene condition (M = 213.5 ms). The differences between Scene vs. Noise and Noise vs. Uniform did not reach statistical signif-

icance, but the difference between Scene and Uniform was statistically significant, $t(11) = 0.04$, $p < 0.04$.

Table 4: Estimated Saccadic Dead Time (ms) in Experiment 2

Participant	Uniform	Scene	Noise
1	74	85	71
2	82	95	82
3	62	72	66
4	81	77	75
5	71	78	72
6	75	90	88
7	72	85	69
8	91	88	73
9	60	67	61
10	60	77	61
11	60	76	69
12	66	76	65
Mean	71	81	71

Saccade amplitudes to single-step targets were also analyzed. They were shortest in the Uniform condition ($M = 6.83°$), followed by the Scene condition ($M = 6.99°$), and the Noise condition ($M = 7.01°$). The difference between the Uniform and Scene conditions was statistically significant, $t(11) = -2.57$, $p = 0.03$, as was the difference between Uniform and Noise, $t(11) = -3.29$, $p = 0.007$. The difference between Scene and Noise was not statistically significant.

1.5 DISCUSSION

The goal of this study was to extend results regarding the programming of saccadic responses to briefly presented targets from a static context to a dynamic context that more closely resembles the environment that saccades are programmed in when viewing naturalistic scenes. In the three conditions of the main experiment (Static, Scene, and Noise), single or double-step targets were presented while participants were engaged in a stable fixation. In the Static task, participants responded to targets following an extended period of fixation

on a central cross. During both the Scene and Noise conditions, participants responded to targets that were presented at the onset of a fixation made while exploring the image. The double-step logic utilized by Becker & Jürgens (1979) was ported to the present study, which allowed comparative inferences to be made regarding the time course of saccade programming in the three conditions.

Modification of saccade programming timelines has been an influential assumption utilized by models that attempt to describe eye-movement control in both scene viewing and in reading (Reichle et al., 1998; Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Previously, these assumptions have primarily been warranted by double-step experiments utilizing tasks similar to the Static condition implemented in our Experiment 1. In the present study, we observed the characteristic logistic shape of the ATF in both static and dynamic viewing conditions, which provides an important confirmation that reprogramming mechanisms that operate in static contexts operate in an analogous manner within a dynamic scene-viewing context.

Amplitude transition functions permit inferences to be made regarding the time course of saccade preparation. SDT represents the point in time prior to the onset of a saccade at which that saccade may no longer be modified by updated visual information. In Experiment 1 we found that when double-step targets are presented in a static movement context in which the targets are presented on a black background, SDT is lower than it is when targets are presented in a dynamic double-step context overlaid on structured backgrounds. This difference in SDT is also complemented by a strong trend towards observing longer SDT in the Scene as compared to the Noise condition, suggesting a possible influence of scene content per se. In Experiment 2, we isolated the influence of background on SDT by comparing three static movement conditions, which differed in the structure of the background stimulus. We found that SDT was no different in the Uniform and Noise conditions but there was a tendency to observe a larger SDT in the Scene condition. Furthermore, by removing movement from the task, SDT estimates in the Scene and

Noise conditions were greatly reduced in comparison to Experiment 1.

In Experiment 1, a notable difference between the Static and Scene task is that within the Scene task, but not the Static task, the targets are presented during a dynamic movement context. A further difference between these tasks is the high-level cognitive processes that are assumed to be activated during scene exploration. In the Scene task, participants are under instructions to view the stimulus to prepare for a later recall phase of the experiment. As a result, in the Scene condition participants are engaged in a more complex task that is more engaging to higher-level cognition. Therefore, it might be predicted that higher order operations specific to the processing of scene elements are implicated in the elevated SDT that was observed in the Scene condition relative to the Static condition. However, a direct comparison between these two conditions does not distinguish between (1) influences that are due to the dynamic movement and (2) differences due to additional high-level scene structure. The Noise condition in Experiment 1 was explicitly included to discriminate between these two alternatives. If scene structure does play a role in determining SDT, then it would be expected that SDT in the Scene condition is also elevated relative to the Noise condition. The pattern of results observed in Experiment 1 suggests that there may be a unique influence of scene content as a trend towards longer SDT was observed for scene backgrounds. SDT was estimated to be 12.98 ms longer in the Scene compared to the Noise condition. While this difference did not reach statistical significance, the pattern of individual results showed a clear trend towards increased SDT in the Scene as compared to the Noise condition. Results from Experiment 2 also support the view that scene content may play a role in determining SDT. As in Experiment 1, a trend for longer SDT was observed in the Scene condition relative to the non-scene conditions. This influence on SDT occurred despite participants receiving instructions that the scene content was not relevant in any way to the successful completion of their task.

The static vs dynamic movement context also appears to play a relatively strong role in determining SDT. In Experiment 1, we observed

a strong effect of movement on SDT. SDT in the two movement conditions (Scene and Noise) was elevated from the Static condition by 22 and 35 ms respectively. As was previously argued, the SDT increase in the Noise condition relative to the Static condition should not be considered to arise from the additional structure present in the $1/f$ amplitude spectra. Experiment 2 provides support for this difference being one that is linked to movement. When movement was controlled for, no difference between the Static (Uniform) and Noise conditions were observed, and the SDT magnitudes in the Scene and Noise conditions were reduced. This was in contrast to Experiment 1 where these backgrounds led to observed differences in SDT. This suggests that differences in SDT can arise when the eyes are in motion and actively engaged in a task, as compared to when they are at rest, waiting for stimulus presentation.

We suggest that the difference in SDT between the Static and Dynamic tasks arises partially due to a form of saccadic pre-preparation that occurs within the static task, but is less likely to be active within the dynamic tasks. We speculate that by presenting the first target, initial stages of saccade preparation may be initiated both to the location of the visible target as well as to the location at which the second target is predicted to appear. In the Static task, the location of the second target is highly predictable as participants are always fixating the same location (central cross) when it appears. In contrast, estimates of the location of the target position in the dynamic task are likely to be far more variable in that there is limited evidence on which to base such predictions. As a result of such predictability we suggest that the saccade motor system may have primed the saccade to the second target prior to the target onset. An alternative possibility is that SDT is elevated in the dynamic tasks due to the occurrence of multiple saccade plan modifications. In this account, when the first target is presented, a modification to the current saccade plan is initiated. Then, upon presentation of the second target, this updated saccade plan must receive additional modification to reach the final target position. It may be the case that such cascaded modifications require the presence of increased SDT. As current models of saccade

programming do not address such a scenario it is difficult to make specific predictions. However, such a simulation would be an interesting and worthwhile exercise.

Stochastic process models of double-step performance in simple tasks provide an elegant framework with which to interpret how these hypothesized differences between tasks could result in SDT variability. Camalier et al. (2007) used a race model architecture to model double-step performance. In this model, saccades are generated as a consequence of a race between a saccade program targeting the first location and an additional stochastic cancellation process that is initiated upon presentation of the second target. The quantity known as the target step reaction time (TSRT) expresses the amount of time that is required by the cancellation process to successfully inhibit saccades to the first target. As lower values of TSRT correspond to lower SDT, the race model predicts that a process that results in a speed up of the cancellation process will also result in a reduction in SDT. The authors applied the race model logic to a double-step task in which the second target step competed with distractor stimuli for saccade target selection. They demonstrated that TSRT was longer when the target and distractors were more similar in color, illustrating that SDT is likely to be longer in conditions where acquisition of the second target is made more difficult. Therefore, it may be the case that lower SDT in the Static task may result from a decrease in the amount of time required for the cancellation process to complete due to the predictable characteristics of the second target location.

Ludwig et al. (2007) measured SDT in two double-step tasks. In the first task, a gap paradigm (Saslow, 1967) was used to observe SDT under conditions in which the latency of responses to targets is systematically varied. Saccade latency was manipulated by completely removing the central fixation cross prior to the onset of the saccadic response. Despite the fact that, on gap trials, saccade latency was considerably reduced (28%), no difference in SDT was observed. In a second static double-step experiment, the authors manipulated the angle of separation between two double-step targets such that the angle between the two targets on a given trial varied between 30° and

90°. In this task, they found a positive relationship between the degree of separation and the estimated value of SDT. They reported a mean SDT of 65 ms at the smallest separation of 30° and a SDT of 113 ms at the largest separation of 120°. The authors introduced a population coding account to explain the pattern of results. In this account, when the initial and final target are presented in close spatial proximity there is shared activation in direction coded movement neurons for the initial and final targets. Therefore, presentation of the initial target provides a form of pre-preparation for the final target movement. Due to such pre-preparation, neurons coding for movement towards the second target have been partially activated and therefore require less time to reach the threshold to initiate a saccadic response. This population coding argument has been shown to adequately account for the effect of target angle separation on SDT (Ludwig et al., 2007). However, the population coding account as formulated in Ludwig et al. (2007) cannot directly account for the results reported here. In all three double-step tasks the angle between the target stimuli was maintained to be a constant 0°. Therefore, according to this framework, activation of neurons coding movement to the second target location by presentation of the first target should be equivalent in both static and dynamic tasks.

The estimated values for SDT observed in our study (Static: 74 ms, Scene: 109 ms, Noise: 96 ms) also provide a measure of empirical grounding to the values for the duration of the nonlabile stage of saccade programming that have previously been suggested (see Table 5). It is important to note that the concept of SDT does not map directly onto that of a nonlabile stage. Firstly, the magnitude of the SDT estimates may vary depending on the method chosen to estimate the point of no return. While a consistent method may be used to compare SDT within a single study, the magnitude of the SDT estimates may vary depending on the amplitude cutoff chosen to calculate SDT. Furthermore, any SDT estimate that is derived from double-step response data includes delays in transmission of the visual information to regions of the brain responsible for saccadic decisions. Since the nonlabile stage of saccade programming is conceptualized to operate

independently of such transmission delays, deriving a nonlabile estimate from SDT requires a subtraction of this afferent delay. Neurophysiological evidence suggests that afferent delays are on the order of 50 ms (for review see Reichle & Reingold, 2013).

Given the current method of calculating SDT, taking such delays into account would suggest that in scene viewing the duration of the nonlabile stage is approximately 60 ms. Due to the task dependent nature of SDT we believe that it is difficult to generalize this result to other domains. We suggest that an intriguing possibility for extension of this work would be to directly compare SDT estimates across both reading and scene-viewing tasks, thereby permitting more concrete inferences regarding the task dependent nature of SDT. Further comment is warranted regarding the efferent delays that are known to exist in the transmission of oculomotor decisions from the brain to the eye. Such delays are thought to be on the order of 20 ms (Becker, 1991). In the saccade programming architecture of the models that we have discussed, such delays must logically occur towards the latter stages of the nonlabile stage. The duration of the efferent delay places a lower bound on the duration of the nonlabile stage as no stimulus information may contribute to the programming of a saccadic response after the movement signal has been sent to the motor effectors.

The present results are particularly informative for models of scene perception (Nuthmann et al., 2010), reading (Reichle et al., 1998; Engbert et al., 2005), and visual search (Trukenbrod & Engbert, 2014) that utilize multi-stage saccade programming architectures. Table 5 shows the parameters used to specify the mean duration of the nonlabile stage of saccade programming in these models. Surprisingly, there is little consistency with regard to the duration of the nonlabile stage within a given task. In some cases, the models predict durations that are less than the minimum interval required by the efferent delay. This represents a potential difficulty in that such models aim to synthesize empirical knowledge regarding the timeline of oculomotor control into predictions about behaviors in high-level tasks. One of the contributions of the present study is to provide a prin-

cipled experimental grounding for assumptions that are critical to such behavioral predictions. However, it is important to note that differences between the double-step tasks under investigation here and other typical scene viewing contexts do exist. For example, in the dynamic viewing tasks, participants are aware that on each trial a sudden target onset will occur after the scene has been explored for some number of fixations. Consequently, it is possible that towards the later stages of a trial participants may begin to modify their viewing strategies. It is therefore worthwhile considering that these tasks are designed to study saccade programming during scene perception tasks that include a target acquisition component.

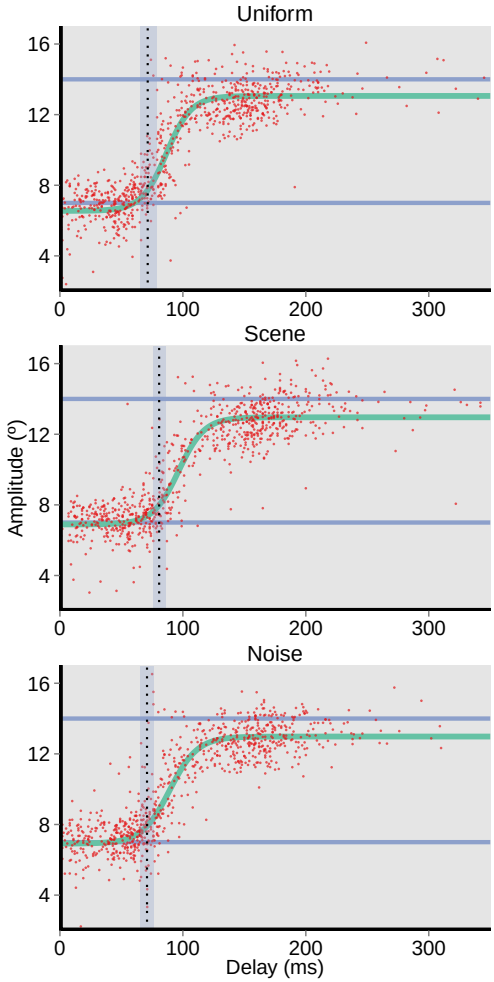

Figure 5: Amplitude Transition Functions for the three experimental conditions in Experiment 2. Points represent the amplitude of saccades initiated in response to the presentation of double-step targets. On the x-axis the delay (D) represents the amount of time elapsed between the presentation of the second target and the onset of the saccade. The y-axis represents the amplitude of the resulting saccade. The green lines represent the best fitting fixed effects curves estimated with a non-linear mixed-effects regression. The horizontal solid blue lines represent the physical distance of the targets in relation to fixation location (7° and 14°). The vertical dotted black lines represent the mean SDT in each condition derived from the by-participant SDT estimates. The light blue bands surrounding the means are bootstrapped confidence intervals.

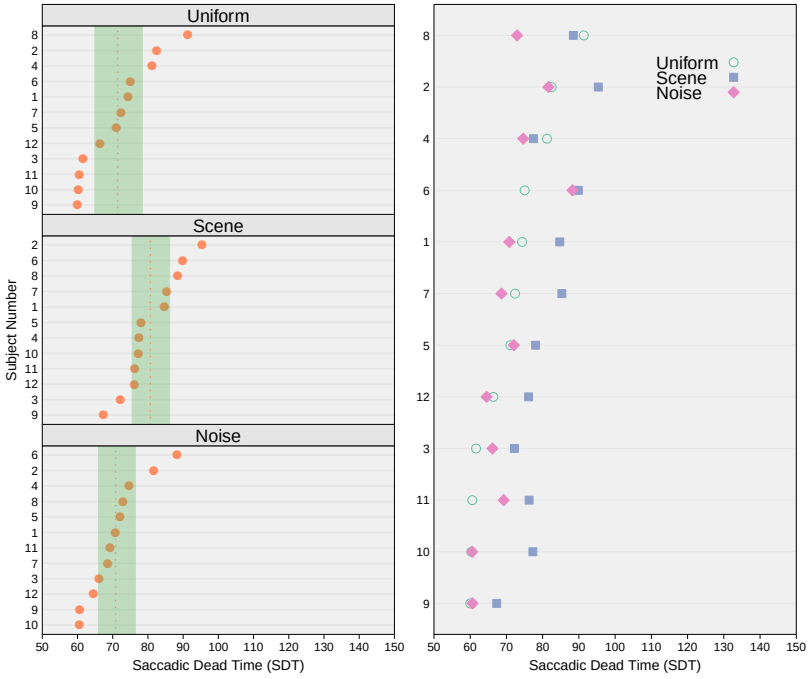

Figure 6: Saccadic Dead Time (SDT) estimates (ms) for the three experimental conditions in Experiment 2. The left plot shows by-participant estimates of SDT. For a given condition, from bottom to top, values are ordered from lowest to highest. The vertical dotted lines represent the mean value of SDT in each condition. The green band surrounding the mean is the bootstrapped confidence intervals of the mean estimate. The right plot shows participants ordered by the magnitude of their estimated SDT value in the Uniform condition.

Table 5: Comparison of model nonlabile estimates

Model			Primary Task	Duration (ms)	Other Task	Duration (ms)
E-Z Reader	1-5	Reichle et al. (1998)	Reading	50	-	-
	9	Pollatsek et al. (2006)		25	-	-
	10	Reichle et al. (2012)		25	Search	25
	10	Reichle et al. (2012)		25	Z-String	25
SWIFT	I	Engbert, Longtin & Kliegl (2002)	Reading	41.6	-	-
	II	Engbert et al. (2005)		[6.1, 58.7][2]	-	-
	II	Nuthmann & Engbert (2009)		[6.1, 58.7]	Z-String	[20.5, 59.6]
	III	Schad & Engbert (2012)		[50, 51.3]	Shuffled	[50, 51.3]
CRISP		Nuthmann et al. (2010)	Scene Viewing	40	-	-
		Nuthmann & Henderson (2012)		40	Reading	14
ICAT		Trukenbrod & Engbert (2014)	Search	40	Reading	40

There is also some discord between models on the question of task-dependent differences in the duration of the nonlabile stage. In both the E-Z Reader model (Reichle et al., 2012) as well as the CRISP model (Nuthmann et al., 2010), there is no mechanism provided by which the mean nonlabile duration may vary within a specific task[3]. In Reichle et al. (2012), an estimated mean duration of 25 ms remained constant both within and across reading, z-string reading and search tasks. In simulations with the CRISP model (Nuthmann & Henderson, 2012), a comparison was made between model predictions of fixation durations in reading and in scene viewing. In these simulations, the duration of the nonlabile stage was permitted to vary across tasks (Reading: 14 ms; Scene Viewing: 40 ms). Similarly to the CRISP model, the SWIFT model also predicts task dependent differences in mean nonlabile durations (Nuthmann & Engbert, 2009). However, as of SWIFT-II, the model additionally assumes that nonlabile durations vary systematically *within* a task by scaling the nonlabile duration to the length of the resulting saccade (Engbert et al., 2005). SWIFT-II was evaluated with data from the Potsdam Sentence Corpus, in which one letter subtends 0.38° and/or 0.45° of visual angle (Nuthmann & Kliegl, 2009). For short saccades ($\approx$ 1°) SWIFT-II predicts average nonlabile durations as long as 58.7 ms. At the other end of the continuum, long saccades ($\approx$ 10°) are predicted to have an average nonlabile duration of 6.1 ms. Therefore, between-task differences in average saccade amplitude may translate into differences in nonlabile durations. However, it should also be noted that estimated parameters in a later version of the SWIFT model result in a model with no between or within-task variability in nonlabile durations (Schad & Engbert, 2012). A recent model, ICAT (Trukenbrod & Engbert, 2014), has modeled fixation durations in both visual search and reading tasks. In ICAT, a fixed nonlabile duration of 40 ms was assumed for both tasks. Future modeling efforts would greatly benefit from addressing

2 As of SWIFT-II, the model predicts a continuum of nonlabile durations. A range of values are provided that reflect short (1°) and long (10°) reading saccades.

3 Although there is no systematic variability built into the nonlabile duration, stochastic variability does enter these models via the duration being sampled from a gamma distribution.

the question of what assumptions are being made with regards to the nonlabile values that are selected, or estimated from the data.

1.6 CONCLUSIONS

Computational models have played an important role in our understanding of oculomotor control behavior in high-level tasks such as scene viewing and reading. However, these models have often relied on simple stimuli used in basic psychophysical paradigms to inform the development of their architectures. Here, using a gaze-contingent double-step paradigm, we have generalized these findings to a more ecologically valid context and revealed important task differences in saccade programming. The present results are particularly informative for the understanding of saccade programming during scene viewing. However, we suggest that future work should directly investigate the processes of saccade cancellation in reading and other task contexts. These studies would provide further generalization for a role of saccade cancellation in eye-movement control, as well as provide empirical validation for task-specific modeling efforts.

Part II

ON THE RELATIONSHIP BETWEEN SCENE QUALITY CHANGES AND FIXATION DURATIONS DURING NATURALISTIC SCENE PERCEPTION.

This part of the thesis addresses the question of how information content acquired during scene viewing determines the length of an individual fixation. Scanning the environment is typically conducted in a goal directed manner. Information that is sampled during an individual fixation is often directly relevant to these goals. For example, in search, the role of a fixation is to identify whether the currently foveated content matches the search target. In a memorization task, foveal processing is required to accurately encoding scene features for later recall. Therefore, one may ask what role the foveated content plays in determining how long we look. When foveal demands are high do we tend to immediately adjust to this difficulty by looking longer? When content becomes easier, do we immediately adapt to look less long? In this part of the thesis I address this question with a series of experiments. The experiments aim to address the basic question: in what ways do fixation durations adapt to processing demands elicited by scene content.

2

EXPERIMENT 1 AND 2 - ON THE IMPACT OF LUMINANCE CHANGES ON THE CONTROL OF EYE-MOVEMENTS WHILE VIEWING SCENES

2.1 INTRODUCTION

The study of eye guidance during naturalistic scene viewing aims to understand the processes that underlie the acquisition of vital visual information from the environment that is relevant to current tasks and goals. Described in a very general manner, investigation into the control of eye movements in scene viewing has proceeded along two primary pathways. The first seeks to address questions relating to where eye movements are directed towards, while the second addresses questions regarding when the eyes move away from currently fixated content. The first question, relating to the spatial aspects of eye movements, has received considerable attention while there is relatively less research investigating the related temporal component (Murray, Fischer & Tatler, 2013). Mean fixation durations in scene viewing are about 300 ms (Rayner, 2009) but there is considerable variability around this mean both within and across individuals. Current understanding of eye-movement programming suggests that some of the variability in the duration of individual fixations may result from factors directly related to oculomotor programming (Becker & Jürgens, 1979; Nuthmann et al., 2010; Walshe & Nuthmann, 2013), as well as global scene properties (e.g. Loftus, 1985; Henderson et al., 2013; Nuthmann et al., 2010), and decisional processes relating to future target selection (Glaholt & Reingold, 2012).

The structure of the mechanisms that govern fixation times has been investigated in a wide variety of tasks (Rayner, 2009). Research that addresses these questions often aims to reveal the manner in which the eye-movement control system adaptively monitors and re-

sponds to environmental demands. A debate of critical importance for the understanding of the temporal characteristics of fixation times is the degree to which stimulus content that is currently under inspection influences the decision of when to terminate the current fixation (Reingold et al., 2012). The hypothesis that fixations are capable of being adjusted on a moment-to-moment basis is referred to as the direct control hypothesis (reading: Rayner & Pollatsek, 1981, scene perception: Henderson & Pierce, 2008; Nuthmann et al., 2010).

This hypothesis is characterised by the assertion that when a fixation duration is under the direct control of stimulus content, there is an immediate adjustment to match the processing demands of the stimulus. In contrast, fixations may be indirectly controlled, and this occurs in the case where fixation times are governed by influences that extend beyond the locally fixated content. For instance, from studies of visual search it is known that fixation durations increase as the complexity of the search array increases (Vlaskamp & Hooge, 2006), when target-distractor similarity is increased (Hooge & Erkelens, 1998; Vlaskamp, Over & Hooge, 2005), and in order to match the difficulty of previously fixated items (Hooge & Erkelens, 1998). These results imply that the eye-movement control system is sensitive, at least in some part, to the global characteristics of the task.

A variety of direct-control mechanisms have been proposed to account for the moment-to-moment adaptation of fixations to current stimulus processing. Concepts related to the structure of direct control mechanisms have seen the most development in theories of fixation times in reading. In reading, a debate exists regarding how the lexical properties of the currently fixated word impacts the time course of that fixation. Mechanisms used to account for such lexical effects may be contrasted as those that implement what is known as a *cognitive trigger*, and those that implement *interference* mechanisms (see Reingold et al., 2012). Cognitive trigger theories postulate that the decision to terminate a fixation is made once the stimulus under inspection has been processed to a sufficient degree, and when this occurs a saccade programme is then triggered. One implementation of such a mechanism is incorporated in the E-Z Reader model, in which

an eye-movement programme is triggered once a superficial stage of lexical processing has been accomplished (Reichle et al., 1998, 2012). In contrast to the triggering mechanisms just described are those that suggest that the variability in the termination of a fixation is a result of difficulties in lexical processing that interfere with the saccade initiation processes. A model that instantiates a variety of direct control along these lines is the SWIFT model (Engbert et al., 2005). In the SWIFT model, the decision to initiate a saccade programme is achieved by an autonomous random timer, and the duration of this timing process may be modulated by the difficulties encountered during lexical processing. Therefore, moment-to-moment difficulties in lexical processing results in increased random timing intervals, and consequently, longer fixation durations.

Although less is known about the mechanisms that govern eye-movement control in scene perception, a model that incorporates an interference mechanism to explain fixation times in this domain is known as the CRISP model (Nuthmann et al., 2010). In this model, an autonomous random walk timer accumulates towards a fixed threshold value and when this threshold is reached, a saccade program is initiated. In the case in which processing difficulties are encountered during scene viewing, the rate at which the timer accumulates to the threshold is reduced. A consequence of such a reduction in the rate of the timer is that the initiation of saccades may be delayed, and therefore longer fixation durations will be observed. An assumption that was made in the original formulation of the CRISP model was that modulations to the timer result exclusively from unidirectional modulations (timer slowdown) (Nuthmann et al., 2010).

An experimental paradigm that has provided some evidence for the direct control of fixations in scene viewing is known as the scene onset delay (SOD) paradigm (Henderson & Pierce, 2008; Henderson & Smith, 2009; Luke, Nuthmann & Henderson, 2013; Nuthmann et al., 2010; Nuthmann & Henderson, 2012). In the SOD paradigm, a scene is masked during a saccade preceding a critical fixation and then restored to full view at varying delays within the critical fixation. Consistently across studies, a population of fixation durations that in-

creased in correspondence with the length of the delay was observed. It was argued that these fixations were increased due to the immediate effects attributable to the missing stimulus. Pannasch, Schulz & Velichkovsky (2011) used a scene based free viewing task in which an irrelevant distractor was introduced either early or late within a critical fixation. Similar to the SOD paradigm, the distractors were presented for variable durations. The results demonstrated that the visual change introduced by the distractor had an immediate prolongation effect on fixation durations, regardless of whether the distractor occurred early or late in fixation, which provided additional support for the direct-control hypothesis.

Going beyond the extreme manipulations of the SOD paradigm, subsequent research has utilised a fixation-contingent scene quality paradigm (Henderson et al., 2013; Glaholt et al., 2013). During selected critical fixations, the entire scene was reduced in quality via a decrease in luminance (Henderson et al., 2013), or by filtering high or low spatial frequencies (Glaholt et al., 2013). Such manipulations are assumed to have deleterious effects on scene processing by influencing the rate at which information is extracted from scenes (Loftus, 1985) as well as impacting the fluent encoding of scene stimuli into working memory (Glaholt et al., 2013). In a study by Henderson et al. (2013), the luminance of the (colour) scene was reduced during the saccade prior to a prespecified critical fixation. During the saccade that terminated the critical fixation, the scene returned to its normal luminance. The durations of the critical fixations were immediately affected by the reduction in scene luminance, with increasing durations for decreasing luminance. Glaholt et al. (2013), on the other hand, demonstrated that fixation durations were affected on a fixation-by-fixation basis depending on the spatial frequency content of the scene stimulus. In their main experiment, during the critical fixation the (greyscale) scene was changed to a high-pass or low-pass spatial frequency filtered version. Under both conditions, fixation durations increased, and low-pass filtering produced a greater effect than high-pass filtering. In a further experiment, the authors additionally modified the orientation of the images, and using a distributional

analysis of fixation durations they were able to differentiate between directly controlled extensions to fixations attributable due to higher-level cognitive influences, and transsaccadic changes resulting in a surprise effect. These results taken together, are highly suggestive that in scene viewing, as in reading, the control of fixation durations is subject to ongoing visual-cognitive processing, such that increases to processing difficulty result in extended fixation durations.

However, further questions regarding the properties of this direct-control process remain. For instance, in the studies that were previously reviewed, the observed effects on fixation durations were primarily ones in which an increase in processing difficulty resulted in an extension to fixation durations. Therefore, these studies demonstrate that there is a tendency for fixations to be immediately adjusted to match the difficulty of the stimulus. However, it is less clear whether the converse is true. That is, will a decrease to fixation durations be observed in the case in which the processing of a stimulus becomes easier and more fluent?

In reading, Kennison & Clifton (1995) investigated the impact of word frequency on two adjacent words embedded in single sentences. High and low word frequency adjectives were followed by high and low word frequency nouns. Parafoveal preview of the noun was prevented by using the invisible boundary technique. When readers first fixated a high-frequency adjective, fixation durations on the subsequent noun showed a word frequency effect, such that longer fixation durations were observed for low-frequency than for high-frequency nouns. In contrast, no such word frequency effect was observed when readers first fixated a low-frequency adjective. Thus, increasing processing demands (high $\rightarrow$ low) resulted in an immediate prolongation of fixation durations, whereas decreasing processing demands (low $\rightarrow$ high) showed no immediate facilitatory effect.

Such an asymmetry in the temporal control of fixation durations has also been observed in visual search. Hooge, Vlaskamp & Over (2007) used a search task in which participants were required to find a closed ring amongst distractor Cs. The distractors in their task varied in the size of the gap, such that small gap Cs were more dif-

ficult to distinguish from the target stimulus than were large gap Cs. They found that fixations on small gap Cs that were preceded by a fixation on a large gap C, showed increased durations. However, a fixation on a large gap C following a fixation on a small gap did not show a corresponding decrease to fixation duration. These results taken together suggest that the control of fixation durations in both reading and visual search tasks involves an asymmetrical pattern of control. While these results provide some guidance on the question of whether asymmetrical control principles generalise to scene viewing tasks, there currently exists no experimental evidence to confirm whether this is the case.

The purpose of the current study was to directly test the hypothesis that the control of fixation durations in scene viewing is asymmetric. To manipulate processing difficulty of the currently fixated stimulus, the present study employed a luminance manipulation such that increased difficulty was obtained by shifting luminance downwards, and decreased difficulty was obtained by shifting luminance upwards. The assumption that modulation of scene luminance levels may be used to control the difficulty of scene processing is derived from several sources. Past research has shown that luminance has strong effects on scene processing, with lowered recognition and recall rates of scenes when they are viewed at a lower level of luminance (Loftus, 1985; van der Linde, Rajashekar, Bovik & Cormack, 2009). These effects are paralleled by an increase in fixation durations to compensate for the increase in processing difficulty encountered due to the luminance reduction (Loftus, 1985). More recently, a control experiment conducted by Henderson et al. (2013) used a free viewing task in which scenes were viewed at 100%, 80%, or 60% original scene luminance throughout the course of the entire trial. They demonstrated that scene luminance had a clear influence on fixation durations such that longer mean fixation durations were observed when scenes were viewed at lower luminance levels. Therefore, these results taken together support the assumption that scene luminance is parametrically related to scene processing difficulty.

In order to test the hypothesis that the direct control mechanism operates in an asymmetric manner, a fixation-contingent scene quality paradigm was used (Henderson et al., 2013). With this method, the luminance shifts took place during saccades when visual transients were suppressed (Ross, Morrone, Goldberg & Burr, 2001; McConkie & Loschky, 2002). While it may be predicted from the gaze-contingent manipulations of Henderson et al. (2013), that longer fixation durations will be observed following a gaze contingent decrease in luminance, it is currently unclear how an increase in luminance will be interpreted by the eye-movement control system during naturalistic scene viewing. The prediction of an asymmetrical direct-control mechanism is that decreased luminance will result in longer fixations, while increased luminance will have no effect. In contrast, symmetrical direct control would predict that shifting luminance down will result in longer fixation durations, and clarifying the stimulus by shifting luminance up will result in shorter fixation durations.

2.2 EXPERIMENTS

2.2.1 *General Methods*

Stimuli

In each of two experiments, participants viewed a total of 100 pictures of real-world scenes, in addition to 4 practice scenes. Each scene had a resolution of 800x600 pixels and was presented in full colour. Scenes were collected from online databases such as google images. They were selected to include a variety of categories such as indoor and outdoor as well as urban and nature scenes. Each scene was viewed by the participant only once over the duration of the experiment, and the experimental scenes were presented to the participants in a randomised order. Initially, scenes were presented at a baseline luminance of 80% in Experiment 1, and 60% in Experiment 2. In order to observe the effect of relative luminance shifts on fixation durations, a luminance transformation was applied. Lumi-

nance shifted stimuli were created by converting the original scene into a $L*a*b$ colour space (Oliva & Schyns, 2000), and modifying the luminance channel L by the appropriate value. This procedure allows the separation of a luminance channel from the two colour channels, and permits the transformation of scene luminance independently of scene colour. Baseline and low luminance conditions for Experiment 1 were constructed by a $L*.8$ and $L*.6$ transformation, respectively. For Experiment 2, a similar procedure was adopted, but the luminance transformation applied was $L*.6$ and $L*.2$. In both Experiment 1 and Experiment 2, the stimulus used in the high (100%) luminance condition was the untransformed scene.

Procedure

Participants were instructed that they would take part in an experiment in which they would see many pictures of naturalistic content and that their task was to encode the scenes for later recall. They were instructed that the recall phase would only begin once all the scenes had been viewed, but were not told how many scenes would be presented. These instructions were provided only to motivate scene encoding behaviour, and therefore the recall phase was not applied. Following the instructions, a nine-point eye-tracker calibration procedure was initiated. A trial began when the participant fixated on a cross presented at the centre of the screen. Following this fixation, the red cross and grey background were replaced with the scene presented at baseline luminance. Participants then engaged in the encoding task until a critical fixation was identified when a participant had made at least 10 saccades since the beginning of the trial. If a critical fixation had been identified, the luminance shift was made during the saccade immediately preceding the critical fixation. The luminance-shifted scene was presented for the entire duration of the critical fixation, and the luminance was then shifted back to baseline during the saccade immediately following the critical fixation. In total, four luminance manipulations were made on each trial; two manipulations resulted in an upward luminance shift, and two manipulations were made in the downward direction. After the first

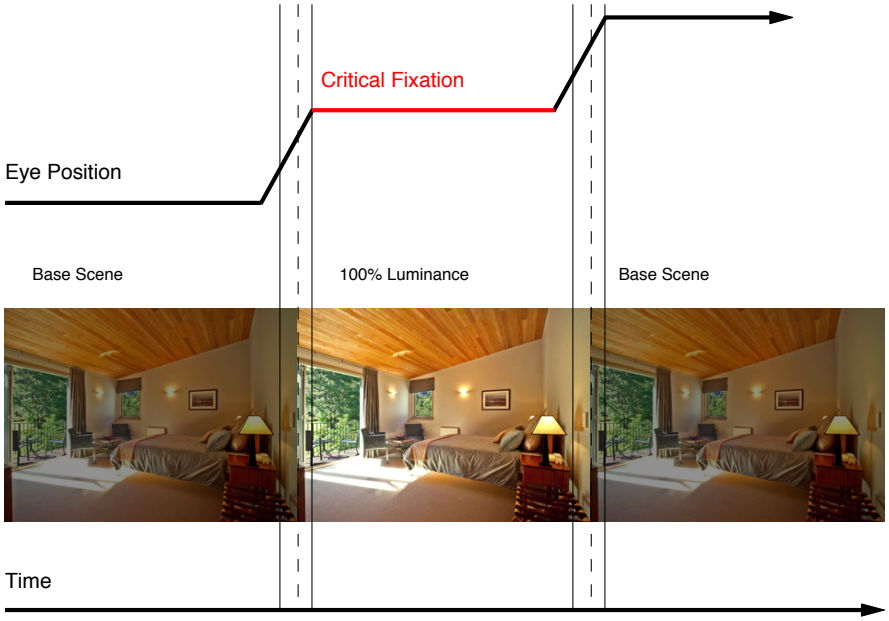

Figure 7: A schematic of the paradigm used to create gaze contingent luminance shifts. Base scenes represent the image that is viewed during the fixation immediately preceding a critical fixation. A critical fixation is defined to occur on the 10^{th} fixation since the previous luminance manipulation. The oblique lines represent saccadic eye movements. During a saccadic eye movement, the scene is either increased or decreased in luminance. A critical fixation is terminated upon detection of a saccadic eye-movement, and the scene is restored to base scene luminance during this saccade.

luminance manipulation had been completed, subsequent shifts occurred on every 10^{th} saccade since the most recent luminance shift. The order of the luminance shift direction (increase vs. decrease), was randomised within a trial. Once the fourth luminance shift had been made, and the participant terminated the resulting critical fixation, one second elapsed until the trial was terminated. The scene was then replaced with a grey background and red fixation cross. Once the participant fixated on the cross, the next trial was initiated. In the situation that the trial lasted longer than 25 seconds, the current trial was abandoned, and the participant was presented with a fixation cross to initiate the next trial. A schematic of the procedure for upward luminance shifts is presented in Figure 1. The mean trial length in Experiment 1 was 18.1 seconds and 19.2 seconds in Experiment 2. The mean number of saccades per trial was 50.3 in Experiment 1, and 48.1 in Experiment 2.

Apparatus

Online detection of saccades involves a speed-accuracy trade off, such that incorporating more samples reduces the noise in the signal. However, by increasing the number of samples, measurement lag is increased, which decreases the temporal precision with which saccades are detected. We implemented a 9-sample online velocity detection algorithm in MATLAB that aimed to mimic Data Viewer's offline saccade detection procedure (SR Research Ltd., 2006). Saccades were identified when gaze data from the right eye reached a two-dimensional velocity threshold of 85°/s. Raw data was post-processed utilising SR Research Data Viewer to parse the gaze samples into sequences of fixations and saccades.

Several data exclusion criteria were applied to remove critical fixations that had been misidentified. Prior to any data exclusion, 97.9% of the luminance manipulations were executed in Experiment 1 and 95.7% were executed in Experiment 2. This number is less than 100%, as a trial would occasionally timeout before all luminance shifts had been completed. Critical fixations on which the display change did not complete prior to fixation onset were discarded. This criteria was validated by comparing the saccades detected online with saccades identified by the post-processed Data Viewer output. Comparison with the post-processed data represents an objective measure, as this data incorporates acceleration and velocity of both prior and future eye-position samples, in detecting current saccadic activity. This resulted in retention of 85.4% of the data in Experiment 1, and 86.4% in Experiment 2. Critical fixations that co-occurred with blinks were also excluded from the analysis. Removing blinks resulted in 67.5% of the critical fixations being retained in Experiment 1, and 68.4% in Experiment 2. A final criteria was applied that excluded critical fixations that had durations of less than 50 ms or greater than 1200 ms, on the assumption that they are not determined by cognitive level processes under investigation in this study (Inhoff & Radach, 1998). As a result of the application of all criteria, 65.8% of the critical fixation were retained in Experiment 1 and 65.1% were retained in Experiment 2.

Analysis

Data were analysed with linear mixed-effects (LME) models, using the `lmer` programme of the `lme4` package (Bates, Maechler & Bolker, 2012) implemented in the R statistical computing software (R Development Core Team, 2012). To evaluate the effect of the downward and upward luminance shifts on fixation duration, we used treatment contrasts in which the baseline condition, where no luminance change occurred, served as the reference group. Consequently, the intercept for the fixed effect "luminance shift", estimates the mean value for the no-shift condition. The two slopes estimate the difference between downward luminance shift and no shift (DOWN) and between upward luminance shift and no shift (UP). The effect of luminance is assessed in the LME model by observing regression coefficients for the luminance shift conditions that are significantly different from 0; a two-tailed criterion of $t = 1.96$ was used to assess statistical significance. The LME models included random intercepts and random slopes for participants and items (Baayen, Davidson & Bates, 2008).

Additional ex-Gaussian distributional analyses of fixation durations were conducted by employing a generalised additive model location, scale and shape (GAMLSS) framework, using the `gamlss` package (Rigby & Stasinopoulos, 2005) implemented in R. GAMLSS is a regression framework that allows the response variability to be modelled by skewed distributions such as the ex-Gaussian distribution. Regression coefficients of the ex-Gaussian parameters contrasted the two treatment conditions (DOWN and UP) with the baseline condition. A two-tailed criterion of $t = 1.96$ was used to assess statistical significance.

2.2.2 *Experiment 1*

Methods

STIMULI The stimuli used in Experiment 1 were presented at a baseline level of 80% of original scene luminance throughout the trial. Upon detection of a saccade preceding the critical fixation, the stim-

ulus was replaced with a scene which had the luminance raised or lowered by a margin of 20%. This meant that in the DOWN condition, participants viewed a stimulus at 60% original luminance, and in the UP condition participants viewed a stimulus at 100% original luminance. During the saccade that terminated the critical fixation, the scene returned to its base luminance.

PARTICIPANTS Four males and 18 females were recruited from the University of Edinburgh student population. The mean age of the participants was 21 years. Each participant was paid £7 per hour of participation in compensation for their time.

Results

The goal of the analysis was to assess the impact that gaze-contingent luminance shifts have on fixation durations. Therefore, our analysis was restricted to critical fixations that began following the termination of a saccade and ended with the initiation of a subsequent saccade. In all cases, the critical fixation was defined such that a luminance manipulation had been made during the saccade immediately preceding the fixation. A baseline measure was constructed in order to detect differences between luminance shifted fixations and fixations in which no luminance shift took place. For each luminance manipulation that survived the exclusion criteria, we measured the duration of the fixation immediately preceding the critical saccade. Since the participant was unaware that a luminance manipulation was to take place during the subsequent saccade, this fixation duration represents an accurate measure of fixation on the unmodified image. It is important to note that a baseline condition with a greater number of observations than were present in either the UP or the DOWN condition was used (cf., Glaholt et al., 2013). A strength of the linear mixed-effects modelling approach adopted in the present study is that it can deal with unbalanced designs (Baayen et al., 2008).

The mean pattern of critical fixation durations is presented in Figure 2. To reiterate, the intercept for the fixed effect of luminance shift estimates the mean value for the no-shift condition ($b = 297.30, SE =$

$9.20, t = 32.30$). As expected, downward luminance shifts were associated with critical fixations that were significantly longer than in the no-shift condition ($b = 44.92, SE = 6.07, t = 7.40$). In addition, there was also a significant increase in fixation durations for upward luminance shifts ($b = 13.28, SE = 4.21, t = 3.15$). The effect of the UP condition is contrary to predictions by both the asymmetric control hypothesis (no change) and symmetric control hypothesis (decrease). When translating the estimated effects of luminance shift into a % increase relative to baseline, it becomes apparent that the effect was much smaller in the UP condition (4.5% increase) than in the DOWN condition (15.1% increase). Comparing the between condition means is informative for the asymmetrical control hypothesis under investigation in the current study. However, changes in mean fixation duration (or the lack thereof) may reflect distinct patterns at the level of underlying distributions. More specifically, previous work on eye guidance in reading and scene perception has argued that applying an ex-Gaussian distributional analysis to fixation durations allows inferences about the time course of effects by quantifying whether effects may be attributed to a shift in central tendency or tail of the distribution (Glaholt & Reingold, 2012; Reingold et al., 2012; Luke et al., 2013; Staub, White, Drieghe, Hollway & Rayner, 2010). The ex-Gaussian is a three-parameter distribution that is derived by a convolution of the Gaussian distribution with the exponential distribution. The parameters contributed by the Gaussian distribution are μ and σ, and describe the central tendency and the spread of the distribution. The τ parameter contributed by the exponential distribution provides a measure of the skewness of the distribution and is useful for describing effects that specifically impact the tail of the distribution.

Figure 3a and c plot the empirical distribution and ex-Gaussian fitted distributions for Experiment 1. Consistent with the findings from the analysis of means, the distributions for both the DOWN and UP condition are shifted to the right relative to the baseline condition, indicating a higher probability of observing longer fixation durations in these conditions. Accordingly, there was a significant effect of DOWN

on μ in the ex-Gaussian fit ($b = 50.06, SE = 2.87, t = 17.40$). There was also a significant effect of UP on μ ($b = 32.04, SE = 2.64, t = 12.11$). A statistically significant effect of DOWN on σ was observed ($b = 10.29, SE = 2.11, t = 4.87$), indicating that the spread of the distribution was larger than in the no-shift control condition. In contrast, there was no effect on σ for the UP condition ($b = 1.59, SE = 1.99, t = 0.79$). An analysis of the τ parameter of the ex-Gaussian fit revealed that the increase in fixation durations in the DOWN and UP conditions does not result from increases that are specifically attributable to the tails of the distributions. Rather, the fixation duration distribution in the UP condition was significantly less skewed than the baseline condition, evidenced by a significant negative effect on τ ($b = -19.92, SE = 3.86, t = -5.13$). DOWN had a small, marginally significant, negative effect on τ ($b = -8.08, SE = 4.14, t = -1.94$).

Discussion

A 20% luminance reduction of the entire scene during critical fixations was associated with an immediate lengthening of those fixations' duration. The pattern of mean fixation durations for a fixation-contingent downward shift in luminance is consistent with results by Henderson et al. (2013). Thus, we provide a replication of their results with a different base luminance level (80% rather than 100%), different scene stimuli and participants, and statistical evaluation that controlled for variability introduced by participants and items. In addition, the current experiment included a condition in which processing was made easier by shifting luminance upwards (from 80% to 100%). There was no facilitatory effect of shortened fixation durations observed in this condition, which is consistent with research in visual search (Hooge et al., 2007) and in reading (Kennison & Clifton, 1995). On the contrary, in the UP condition we observed a significant lengthening of fixation durations, but the magnitude of the increase to fixation durations was considerably smaller than in the DOWN condition. Taken together, the results are indicative of an asymmetrical pattern of control such that difficulties in scene processing are directly incorporated and result in longer fixation durations, whereas

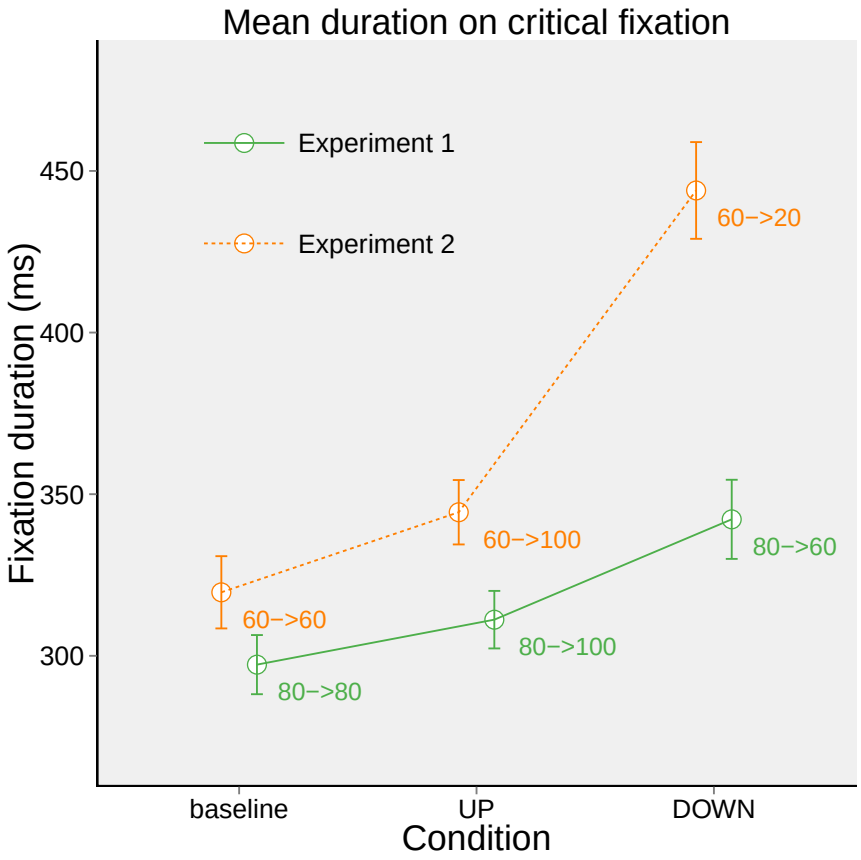

Figure 8: Mean fixation durations on critical fixations following gaze-contingent luminance shifts. Fixation durations are plotted as a function of the direction of luminance shift. Data is plotted for Experiment 1 (solid line) and for Experiment 2 (dashed line). Error bars represent the standard error of the mean.

processing facilitation does not lead to a comparable decrease in fixation durations.

One possibility for the lack of a speedup in the UP condition is that the magnitude of the luminance difference between the baseline and increase in luminance was insufficient to provide enough processing facilitation to elicit shorter fixation durations. That is, the possibility remains that while a luminance shift from 80% to 60% is sufficient to create scene processing difficulties, a shift from 80% to 100% is insufficient to create a context for processing facilitation. This hypothesis is strengthened by the results of the distributional analyses. This analysis showed that in the UP condition, an overall shift in the distribution occurred due to a significant positive effect on μ. However, we also observed a significant negative influence on the tail of

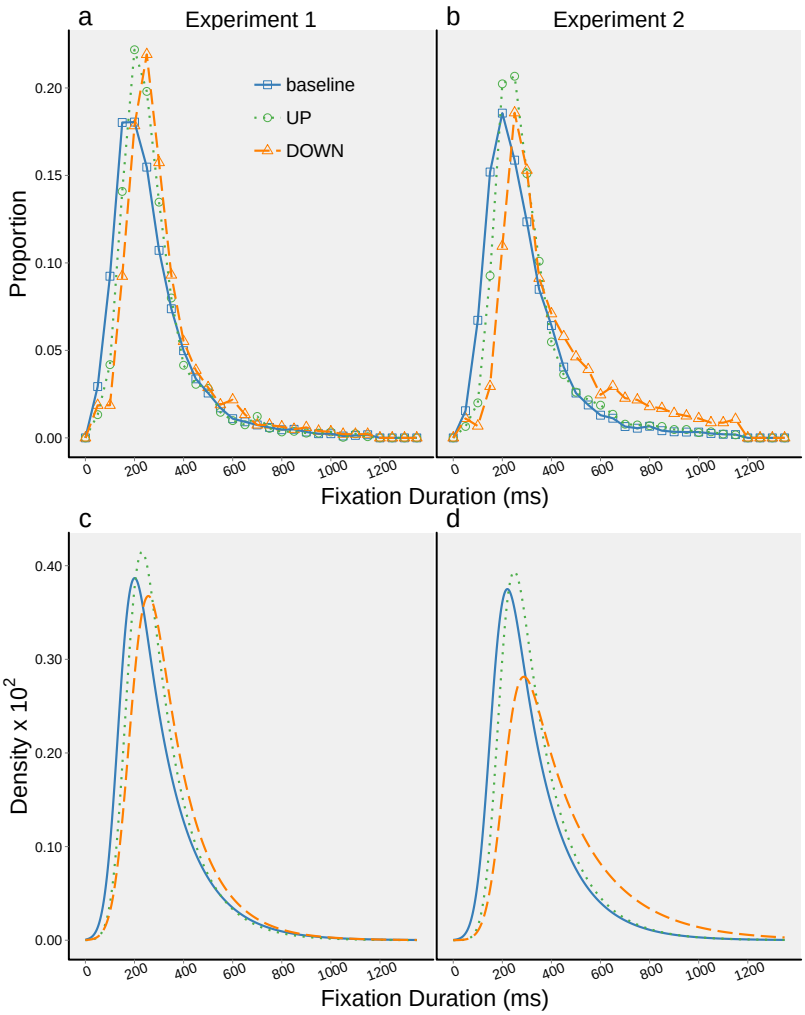

Figure 9: Fixation duration distributions. Empirical distributions for the three luminance conditions in Experiments 1 (Panel a) and Experiment 2 (Panel b), and their respective ex-Gaussian fitted distributions plotted in (Panel c) and (Panel d).

the distribution (decrease in τ), indicating a significantly less skewed distribution in the UP condition. Therefore, we hypothesise that a more extreme luminance enhancement may result in a diminished impact on the central tendency of the distribution than was observed in Experiment 1, but that the influence on the tail of the distribution will remain. Experiment 2 was designed to address this possibility by lowering the baseline luminance of the scene to 60% and further lowering the luminance to 20% in the DOWN condition and raising it to 100% in the UP condition.

2.2.3 *Experiment 2*

Methods

PROCEDURE AND STIMULI The procedure and stimuli for Experiment 2 were identical to that of Experiment 1 in all aspects other than the magnitude of the luminance change. During the saccade immediately preceding a critical fixation, the luminance was either shifted up to 100% or down to 20% luminance, from a 60% luminance baseline. During the saccade immediately following the critical fixation, the luminance of the scene was changed back to the 60% baseline level.

PARTICIPANTS 13 females and 4 males who did not participate in Experiment 1 were tested in Experiment 2. The mean age of the participants was 24 years. Each participant was paid £7 per hour of participation in compensation for their time.

Results

Experiment 2 sought to complement the results observed in Experiment 1 by testing whether similar effects would be observed when a different baseline luminance level was used, and when the magnitude of the luminance shifts was increased. The observed pattern of mean durations is plotted in Figure 2. In the LME model, the intercept for the fixed effect of luminance shift estimates the mean value for the no-shift condition ($b = 319.47, SE = 11.09, t = 28.79$). Experiment 2 used a lower baseline level of original scene luminance than Experiment 1 (60% vs. 80%). Accordingly, the mean fixation duration in the no-shift baseline luminance condition was longer in Experiment 2 than in Experiment 1 (319 ms vs. 297 ms, Figure 2). Following the default prediction, downward luminance shifts were associated with critical fixations that were significantly longer than in the no-shift condition ($b = 124.28, SE = 13.15, t = 9.44$). Experiment 2 used a greater magnitude of luminance shifts than Experiment 1 (40% as opposed to 20%). Therefore, downward shifts in luminance resulted in a larger relative increase in fixation duration in Experiment 2 as

compared to Experiment 1 (Figure 2). In addition, there was again a significant increase in fixation durations for upward luminance shifts ($b = 24.55, SE = 7.92, t = 3.10$). Relative to the no-shift baseline condition, fixation durations increased by 38.9% in the DOWN condition but only 7.7% in the UP condition.

The approach to analysing the distributional effects in Experiment 2 was conducted along analogous lines to Experiment 1. Figure 3b and d show the empirical and ex-Gaussian fitted distributions. A similar pattern was found to Experiment 1 in that the distributions showed a general rightward shift consistent with the increased mean durations observed in both luminance shift conditions. The GAMLSS model yielded a significant positive effect on μ for both the DOWN and UP conditions relative to the no-shift baseline condition (DOWN: $b = 51.63, SE = 3.65, t = 14.13$; UP: $b = 31.11, SE = 3.02, t = 10.28$). In the DOWN condition, there was a significant positive effect on σ ($b = 7.42, SE = 2.71, t = 2.73$), indicative of an increase in the variance in this condition. As in Experiment 1, there was no effect on σ in the UP condition ($b = -2.77, SE = 2.28, t = -1.21$). With regard to the τ parameter, a different pattern of results was observed than in Experiment 1. In the DOWN condition, there was a substantial increase in long fixation durations, which is manifested as a more positively skewed distribution. This late influence on the tail of the distribution was substantiated by a statistically significant positive effect of DOWN on the τ parameter ($b = 70.14, SE = 6.32, t = 11.08$). No statistically significant effect of the UP condition on τ was observed ($b = -6.51, SE = 4.62, t = -1.40$).

Discussion

A possible explanation for the observation that no facilitatory effect was observed in Experiment 1 is that the magnitude of the luminance increase was too small to result in benefits in processing to the degree required in order to observe shortened fixation durations. Experiment 2 directly tested this hypothesis by increasing the magnitude of the luminance shift from baseline in both the UP and DOWN condition. Mean fixation durations observed in Experiment 2 showed a

similar pattern to Experiment 1. Further decreasing the luminance of the scene during selected critical fixations was associated with an immediate and substantial increase in fixation duration. Furthermore, we did not observe a decrease in fixation durations following a facilitation in stimulus processing, as was assumed to occur following the increase in scene luminance. By inspecting the parameters of the ex-Gaussian distribution in Experiment 1, we speculated that if the more extreme luminance shift in the UP direction diminished the influence on the central tendency of the distribution then a facilitation effect may have been observed. The results from the analysis of means and parameter of the ex-Gaussian fit suggest that this is not the case. These results complement Experiment 1 and provide further support for the hypothesis that fixation durations are controlled in an asymmetric manner. The results from Experiment 1 and 2 show that a fixation-contingent increase of overall scene luminance was not sufficient to elicit a speedup in processing as observed through decreased fixation durations.

2.3 GENERAL DISCUSSION

Two experiments were conducted to test whether the adjustment of fixation durations in naturalistic scene viewing is unidirectional (slow down), or bidirectional (speed up and slow down). A saccade-contingent display change method was used to make the scene more difficult or easier to process during prespecified critical fixations. In Experiment 1, a luminance baseline of 80% was presented to participants and the luminance was shifted to either 60% (DOWN) or 100% (UP). Experiment 2 extended these results by reporting a similar pattern for a 60% baseline with shifts to 20% (DOWN) and 100% (UP). If the direct-control process was asymmetric or unidirectional, decreasing the luminance of the scene should make processing more difficult and result in longer fixations, while clarifying the scene by increasing the luminance should have no effect on the duration of critical fixations. In contrast, if fixation durations were controlled in a symmetric or bidirectional manner, shifting luminance down should result in

longer fixation durations, and shifting luminance up should result in shorter fixation durations. In both experiments, a pattern consistent with the asymmetrical hypothesis was observed such that decreases to luminance resulted in longer fixation durations, but increases to luminance did not result in an immediate decrease in fixation durations.

Downward luminance shifts were associated with increases in fixation durations in both Experiment 1 and Experiment 2. This was reflected in a difference in elevated mean durations relative to the baseline luminance. The overall effect of decreasing luminance on fixation durations is broadly a replication of results reported by Henderson et al. (2013) with different baseline conditions (60% and 80% compared to 100%) and novel stimuli. Additional distributional analyses using GAMLSS regression models qualified the time course of the observed effects. The results from the distributional analysis for Experiment 1 revealed that the increased durations in the DOWN condition occurred due to an overall shift in the distribution (increase in μ) as well as a significant increase in σ, the latter indicating the presence of greater variability in fixation durations in this condition. By comparison, the comparatively larger increase in durations in Experiment 2 was again associated with an overall shift in the distribution (increase in μ) and an increase in σ, but also with a longer tail (increase in τ). The specific influence on the tail of the distribution in Experiment 2 may be partially informed by a recent study conducted by Glaholt et al. (2013). In their study, the authors used a fixation-contingent scene quality paradigm to modify scenes under a variety of conditions such as spatial frequency filtering, and changes to the orientation of the image. In order to observe the differential effects of these modifications on fixation durations, they reported ex-Gaussian fitted distributions for the various conditions. They found that effects on the tail of the distributions were observed primarily for conditions in which the manipulation was hypothesised to result in a change that presented challenges to the later stages of stimulus encoding. In Experiment 2 of the current study, the extreme luminance manipulation (60% $\rightarrow$ 20%) is likely to lead to difficulties in integrat-

ing the lowered-luminance stimulus into existing working memory structures and may partially account for the overall increase in fixation durations and the effects observed on the tail of the distribution.

Upward luminance shifts were associated with a small but reliable increase in fixation durations, which is contrary to predictions by both the asymmetric control hypothesis (no change) and symmetric control hypothesis (decrease). The distributions revealed that in both experiments the increase was attributable to an increase in the central tendency (increased μ); there was no increase in τ in either experiment (rather a significant negative effect in Experiment 1), and no effect of UP on σ.

One explanation for this small increase comes from an effect of surprise that may accompany the shift of luminance that participants encounter on critical fixations. The analysis provided by Glaholt et al. (2013) is informative of why this might be the case. They found that fixation durations were increased in all conditions, but that the effects on the tail were absent for the conditions in which no encoding difficulty was to be expected. These contrasting effects were explained by suggesting that the fast-acting effect that influences the central tendency is a result of surprise due to a detected mismatch between transsaccadic stimulus content. The small but significant increase in fixation durations in the UP condition of both experiments reported here is consistent with the fast-acting effect of surprise that is hypothesised to occur following transsaccadic changes to the scene. Their study included another control experiment that is relevant to the interpretation of the present results. During critical fixations, colour information was added to the greyscale scene. By clarifying the stimulus with a colour enhancement, stimulus processing should be facilitated. According to the symmetric control hypothesis, adding colour should lead to an immediate decrease in fixation duration. However, an increase in the durations of critical fixations was observed, which resulted from an increase to μ, but not from τ. These results are consistent with the results reported here.

Our presentation of the distributional effects that further qualify the inferences made by assessing differences in mean fixation du-

rations is in keeping with recent analyses in reading (e.g., Glaholt, Rayner & Reingold, 2014; Luke et al., 2013; Reingold et al., 2012; Staub et al., 2010) and scene viewing (Glaholt & Reingold, 2012; Glaholt et al., 2013; Luke et al., 2013). Such analyses are highly informative in that they reveal the specific components of the distributions that contribute to the observed mean effects. As has been previously discussed, these results contribute to a growing body of research demonstrating consistent distributional effects within a variety of viewing tasks.

The pattern emerging from the present study, as well as recent empirical results, is that the direct control mechanism operates in an asymmetric manner, in both scene viewing and other visual-cognitive tasks. For instance, Glaholt et al. (2014) reported an asymmetrical control pattern in a reading task in which the contrast of the sentence text was either increased or decreased in a gaze-contingent manner. During the saccade immediately preceding a critical fixation the contrast of the sentence text with the background was either increased, decreased, or was left unchanged. The authors found that upon landing on a sentence that had decreased contrast, fixation durations were increased relative to the no change baseline condition, whereas fixation durations remained the same when contrast was increased. Such results complement previous results observed in both reading (Kennison & Clifton, 1995) and in visual search (Hooge et al., 2007).

The results reported here have direct theoretical consequences for models of eye-movement control generally, but most specifically for accounts of fixation behaviours in scene perception. A computational framework that has had considerable success in modelling the temporal aspects of eye-movement control in scene viewing is known as the CRISP model (Nuthmann et al., 2010; Nuthmann & Henderson, 2012). The CRISP model is a stochastic timing model such that a random walk timing process accumulates to a fixed threshold value. Once this threshold is reached, the programming of a saccade is initiated. The variability of fixation durations predicted from the model are generated from three primary sources, a) the inherent stochasticity of the random walk timer, b) modulation of the random walk's

transition rate due to difficulties encountered during stimulus processing, and c) cancellation and reprogramming of current saccade programmes. In the original formulation of the CRISP model it has been assumed that eye-movement control operates in a manner consistent with what we have here called asymmetric control. That is, modulations to the timer could only occur due to processing difficulty that is expressed as a timer slowdown. With regards to the present results, the CRISP model captures such behaviour by assuming that difficulties in processing due to the decrease in luminance, result in a slowdown of the random walk timer rate and a temporal increase in the interval between successive saccade programmes. However, the results reported here with respect to the condition in which luminance is increased suggest that the default timer slowdown implemented in the CRISP model is sufficient to capture the effects of both degrading and enhancing stimulus processing.

A relevant question for future studies is how the adaptation of fixation durations to immediate changes in processing difficulty changes over the course of viewing. One possibility is that fixation durations may adapt with an immediate increase when processing difficulty increases but may decrease more gradually, say on the second or third fixation, following a decrease in difficulty. Trukenbrod & Engbert (2014) have investigated this issue using a task that required participants to scan sequences of horizontally arranged symbols from left to the right to search for a target stimulus. The target was always a ring, and Landolt-Cs were used as distractors. Processing difficulty of the stimulus elements was manipulated by increasing or decreasing the size of the gap in the Landolt-Cs. Fixations durations upon first encountering a change of difficulty, as well as fixation times on subsequent elements were measured. The authors reported an asymmetrical pattern of control of fixation durations such that increasing difficulty resulted in an increase to fixation durations upon first encountering the change, while a decrease in difficulty resulted in no immediate impact. However, they reported a delayed adjustment to fixation durations in the decreasing difficulty condition, as fixation durations showed evidence of a decrease for later fixations. The time

course of the adjustment to changes in processing difficulty within scene viewing is currently an open empirical question.

2.4 CONCLUSION

In summary, this study used a luminance manipulation in order to vary the scene processing difficulty in a gaze-contingent fashion on critical fixations. We predicted that if the control of fixation durations operates in a symmetric manner, then shifting luminance down would result in increased fixation durations, while shifting luminance up would result in decreased fixation durations. On the other hand, if control is asymmetric we predicted that decreasing the luminance would result in fixation duration increases and that luminance increases would result in no change to fixation durations. The pattern of results observed in the two experiments provides support for the asymmetric control hypothesis.

3

ON THE SEARCH FOR FACILITATION

This is a very complicated case, Maude. You know, a lotta ins, lotta outs, lotta what-have-you's. — Jeff "the Dude" Lebowski

3.1 EXPERIMENT 3 - ON THE IMPACT OF CHANGES TO THE QUALITY OF CENTRAL VISION

3.1.1 *Introduction*

In Experiment 3 we address the question of how the mechanisms responsible for the adaptive control of fixation durations respond to scene quality changes that occur only within central vision. Previous research has shown that fixation durations increase when the quality of a scene is severely reduced within the central region of the visual field (Nuthmann, 2014; Laubrock, Cajar & Engbert, 2013).

The goal of this experiment was to assess whether viewing the scene in degraded form in the central visual region only would result in decreased fixation durations when the central patch was gaze contingently increased to match the luminance of the peripheral region. We included a central region reduction in luminance to make the change unpredictable and therefore comparable to the manipulations made in Experiment 1 and Experiment 2. The motivation for this experiment is derived from the speculation that peripheral preview of the upcoming saccade target location may be having a substantial influence on governing fixation durations upon landing at that location. In Experiments 1 and 2 it is always the case the the luminance of the region corresponding to the critical fixation is always previewed in the reduced, baseline luminance. Therefore, if this peripheral preview is having an influence on the timing of the upcoming eye-movement it may be that this is masking the influence of any

facilitatory effect due to the increase in scene luminance. Therefore, to test this hypothesis we designed a scene viewing task in which the luminance of the periphery was presented at the same original scene luminance throughout the entire viewing session. However, we used a moving mask to degrade or enhance the central visual region during critical fixations and observed the impact of this manipulation on fixation durations.

3.1.2 *Methods*

In Experiment 3, 23 participants were tested and included in the analysis. Participants were recruited from the University of Edinburgh student population and were paid £6 an hour in compensation for their time. The equipment used in this Experiment was the same as that of Experiment 1 & 2. The trial structure and images were the same as in Experiment 1 & 2. Experiment 3 differed from the previous experiments in that a gaze contingent mask was presented to participants as they engaged in the recognition task. Participants were informed of the presence of the mask prior to the experiment. The mask was presented at a baseline luminance of 60% and the region external to the mask was presented at original scene luminance (100%). The masked region was updated to be in direct correspondence with the centre of gaze that is reported by the eye-tracker. Human foveal vision, the region with highest acuity, encompasses the central 2 degrees of vision while central vision refers to the part of the visual field subtending 5 degrees of visual angle. Any part of the visual field outside this region is referred to as peripheral vision. The mask had a diameter of 4°. Therefore, the mask fully encompassed foveal vision but was slightly less than central vision. The luminance filtered scenes were created offline and creating the mask patches did not consume processing resources or cause delays in the presentation of the stimuli to the participants. Filtering was implemented in MATLAB2012a using an algorithm from Nuthmann (2014). The experiment was implemented in MATLAB2009b using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner et al., 2007).

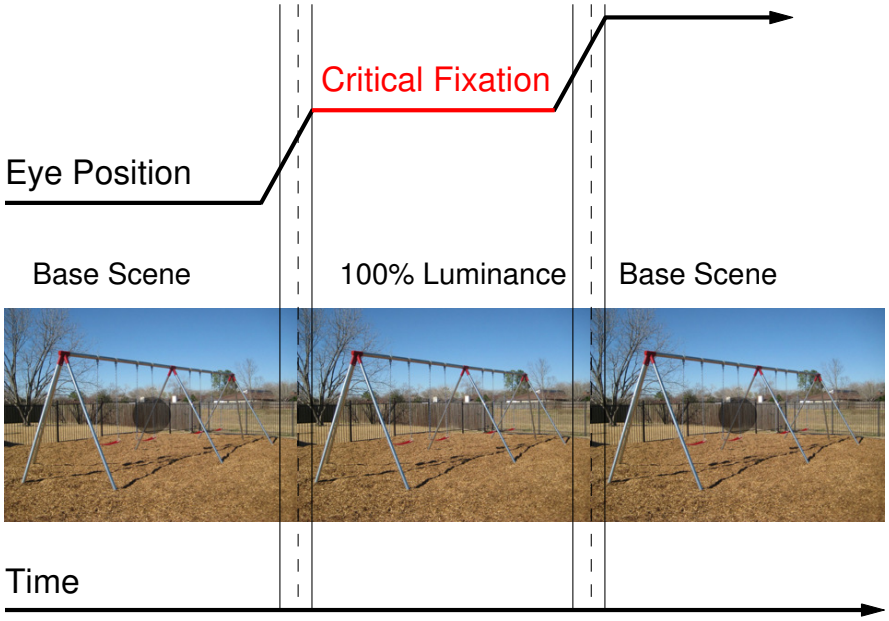

Figure 10: Design Figure. The scene is viewed through a mask that is presented at the central visual region. The mask is always presented at 60% of the normal scene luminance. On a critical fixation the luminance is shifted up or down. When shifted up, the luminance of the masked region now completely matches the background. Note that the background is always presented at normal scene luminance. When luminance is shifted down it is shifted to 20% normal scene luminance. Note that this scene and mask are intended for illustrative purposes only, the scale and filter are not exactly those that were used in the experiment.

The luminance manipulation conditions were defined in an identical manner to the previous experiment such that each trial consisted of 4 luminance shifts, two luminance increases and two luminance decreases. Upon detection of a critical saccade the mask luminance was either shifted down to 20% original scene luminance or up to 100% scene luminance. No change was made to the luminance of the region surrounding the mask. In the UP condition, the mask was shifted to exactly match the luminance of the region surrounding the mask.

3.1.3 *Results*

The results from Experiment 3 did not demonstrate a speed-up in fixation durations within the the UP condition as had been hypothesized

to occur. The analysis approach was similar to Experiment 1 & 2. A linear mixed effects regression with random intercepts and slopes varying in a by-subject manner were used to quantify the influence that modifying the mask had on fixation durations. The estimate for the baseline FDs were, $b = 328.759, SE = 10.02, t = 32.08$. A significant effect of UP was observed, $b = 16.23, SE = 7.83, t = 2.07$ while no statistically significant effect of condition DOWN was observed, $b = 1.13, SE = 4.90, t = 0.23$. We did not analyze distributions as was done in Experiment 1 & 2 as the primary purpose of this experiment was to investigate whether decreased fixation durations would be observed following an increase in scene information. Since this was not observed in the mean responses we did not proceed to the more specific distributional analysis.

3.1.4 *Discussion*

The purpose of Experiment 3 was to test whether a speed-up of fixation durations would be observed in the case that the luminance of a gaze contingent moving mask was increased to match the luminance of the periphery. We conducted this test to observe the influence on fixation durations when trans-saccadic increases to the scene stimulus quality of central visual regions are made. Similarly to the Experiment 1 & 2, we observed that an increase in scene quality resulted in a small but significant increase in fixation durations.

A comment is warranted regarding the observed effects in the DOWN condition of Experiment 3. We included this condition in order to balance the design and ensure that the design was comparable to the previous two experiments. In the DOWN condition an unanticipated observation was made. Even though central vision was degraded and therefore made more difficult to process, fixation durations showed no increase relative to the baseline. Based on Experiments 1 & 2 as well as previous reports in the literature (Walshe & Nuthmann, 2014; Henderson et al., 2013; Loftus, 1985) it would be anticipated that a degradation of this region would lead to an increase in fixation durations. However, we speculate that the reason that no FD increase

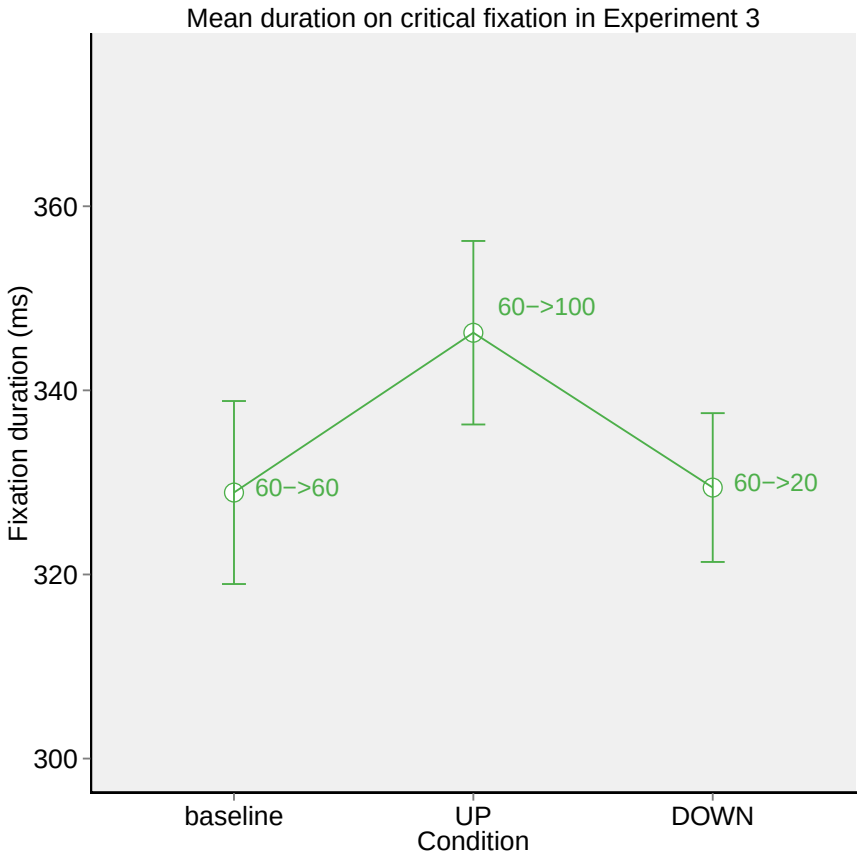

Figure 11: Experiment 3 Means. The mean fixation durations for the three experimental conditions are presented. The measurement in the baseline condition represents a fixation duration that is observed with the central mask at 60% luminance and the peripheral luminance at 100%. The measurement in the UP condition is measured on critical fixations when the central mask is raised to 100% luminance. In this case, the central mask matches the peripheral luminance level. In the DOWN condition, the central mask is lowered to 20% luminance.

is observed is due to the relative luminance levels of the central and peripheral regions. In this experiment when a critical fixation begins in the DOWN condition the luminance has been lowered to 20% original luminance. This level is so low that nearly no scene content can be extracted from the central region. However, luminance in the periphery is still presented in high quality. A possibility is that the relative difference between the high (peripheral) and low (central) visual information results in the rapid generation of a movement signal. A study conducted by (Henderson, McClure, Pierce & Schrock, 1997)

provides some support for this hypothesis. In this study, isolated objects were placed horizontally in a row and participants were required to fixate each object in order from left to right. The task was to encode the objects for a subsequent recall task. Furthermore, in one condition participants viewed the stimuli with a moving mask that was located at the centre of their gaze and removed access to central vision. The authors found that the gaze duration (sum of first pass fixation durations) on the objects in the mask present condition was shorter than in the mask absent condition. The authors discuss the possibility that their results may arise due to the visual system rapidly moving gaze (and attention) away from the central region where the scene quality is low, to the peripheral regions where scene quality is high. This hypothesis is not addressed further within the thesis. However, future work could attempt to clarify these results by varying the degree of mismatch between central and peripheral visual content. For instance, what happens when the central reduction is not so severe? A prediction could be that when the central degradation is less severe that fixation durations may show an increase as the visual system now attempts to extract hard to process information from this region.

3.2 EXPERIMENT 4 - ON THE DELAYED ADJUSTMENT OF FIXATION DURATIONS TO GAZE CONTINGENT LUMINANCE CHANGES.

3.2.1 *Introduction*

So far, our investigations along with recently reported results in the literature (Glaholt et al., 2013) suggest that fixation durations do not tend to decrease in duration following gaze contingent decreases in scene processing difficulty (see Chapter 2). In Chapter 1 we argued that the immediate and rapid increase in fixation durations that occurs following a change in scene luminance is a result of a process that we call surprise. The source of this surprise may be a mismatch between the expectations of the pre- and post-saccadic stimulus content, or it could be that this surprise is a result of the visual system rapidly adapting to a new and unexpected level of luminance that is encountered upon arriving at the post-saccadic location.

In the following study, an experiment was designed that would allow us to explore the question of how fixation durations adapt over time to changes in the difficulty of processing of scene stimulus content. When the luminance of the entire scene is changed and remains changed, what pattern of adjustment will be observed for the subsequent fixations? In other words, we no longer restrict our analysis to the n + 1^{th} fixation as we did in Chapter 2 but extend our analysis to the sequence of fixations that follow the change. From our previous studies, it was reasoned while the first post-change fixation would be influenced by surprise, subsequent fixations would be relatively less influenced by such a surprise effect. The original aim was to observe the impact of quality changes on these non-surprise influenced fixation durations. For instance, we know from classic work on the impact of luminance in scenes that increased luminance leads to decreased fixation durations (Loftus, 1985). However, from Chapter 2 we predicted that this decrease does not happen immediately following a change in luminance, primarily because of the inhibitory surprise effect. Therefore, it was reasoned that the adaption towards

decreased fixation happens gradually over viewing. The design in the present experiment explicit tests this delayed adaption hypothesis.

3.2.2 *Methods, Equipment and Materials*

The equipment used in this experiment was identical to those studies reported in Chapter 2. The online detection of saccade was also conducted in an identical manner (see Section 2.2.1). The experiment procedure was also identical to the procedure as described in Section 2.2.1 except for the following details. Participants viewed the scene at a lower luminance of 60% original scene luminance for 10 fixations. Following 10 fixations the scene was increased in luminance. The luminance was increased by 40% to the original level of scene luminance. Participants then viewed the scene at the new luminance level for 10 fixations. After these 10 fixations the scene was then lowered in luminance back down to 60%. This pattern was repeated once more and then the trial was terminated. In total, participants were presented with four luminance changes per trials. Therefore, there were two types of luminance changes in this experiment; (60% $\rightarrow$ 100%) and (100% $\rightarrow$ 60%).

An important difference to note about this experiment as compared to the previous experiments is that the participants could anticipate the direction of the luminance change that would be presented. For example, if the luminance was currently at the baseline level the luminance shift was always an increase. Likewise, if the luminance was high the shift was always back to baseline. Therefore, a participant could easily extract this pattern to have knowledge of the direction of the upcoming luminance shift.

3.2.3 *Results*

The primary aim of this analysis was to investigate the modulation of fixation durations to changes in scene luminance as a function of the ordinal number of the fixation relative to the change and the direction

of the change (i.e. upward luminance shift or downward luminance shift). The models included random intercepts and random slopes that varied by subject. Two separate models were tested, one for data in which the direction of the shift was upwards and another for the data which the shift was downward. The coefficients from these models are presented in Table 6. The coefficient for the intercept represents an estimate of the mean fixation duration for fixations that occurred prior to any shift (i.e. a baseline fixation duration measurement). In the model of the upward luminance shift this is an estimate of the mean fixation duration on scenes that had a luminance of 60%. A dummy code regression model was used to assess what influence the order of the fixation number after a shift had on modulating fixation durations from the baseline level. Therefore, the coefficients for fixation 1...5 indicate the change in mean fixation duration for that fixation number relative to the baseline fixation duration. We consider a reliable difference to be observed when ($|t| > 2$)

The analysis revealed an unexpected result. It was observed that in the UP condition there is an immediate decrease in fixation durations that immediately follow the luminance change (i.e. Fixation 1) ($b = -21.14, SE = 5.77, t = -3.66$). Fixations > 1 also tended to be shorter and by the 5^th post change fixation fixation durations were estimated to be approximately 18 ms shorter than baseline fixation durations ($b = -18.54, SE = 5.45, t = -3.40$). The relationship between ordinal fixation and fixation duration is graphically demonstrated in Figure 12.

In the downward luminance shift condition we observe the characteristic increase in fixation durations on the first post change fixation. Shifting luminance down results in an immediate increase (i.e. Fixation 1) in fixation durations ($b = 25.20, SE = 4.82, t = 5.22$). On Fixation 2 there is a drastic reduction in fixation durations such that the average duration is brought below the average baseline fixation duration ($b = -11.76, SE = 5.12, t = -2.30$). For subsequent fixation durations the expected pattern of increased fixation durations is again observed and by the 5^{th} fixation fixation durations are estimated to be approximately 19 ms longer than in the baseline

($b = 19.43, SE = 5.29, t = 3.67$). The overall pattern of results in the DOWN condition is visualized in Figure 12.

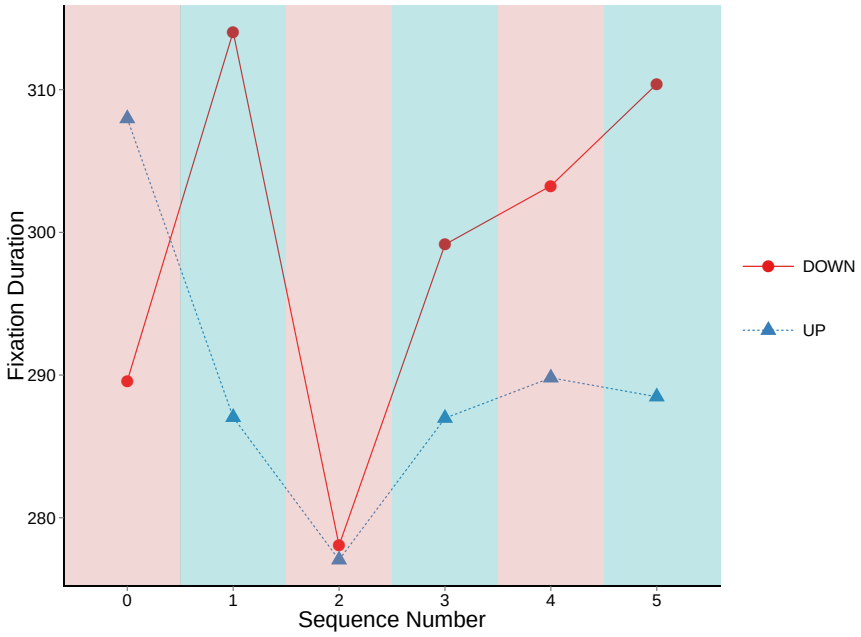

Figure 12

	DOWN			UP		
	Estimate	Std. Error	t value	Estimate	Std. Error	t value
Intercept	289.80	8.69	33.36	308.65	10.65	28.97
Fixation 1	25.20	4.82	5.22	-21.14	5.77	-3.66
Fixation 2	-11.76	5.12	-2.30	-33.12	5.19	-6.38
Fixation 3	9.58	5.25	1.82	-22.43	4.76	-4.71
Fixation 4	13.14	5.73	2.29	-19.82	4.75	-4.18
Fixation 5	19.43	5.29	3.67	-18.54	5.45	-3.40

Table 6: Dummy coded regression tests for long range dependencies. The coefficients from the linear mixed-effects regression are provided. Two regression models were fit, the two models are distinguished by the headings *Up* and *Down*. The models were fit to data depending on whether the observation was derived from a luminance increase, or decrease. The intercept estimates the mean fixation duration for the baseline measurement. The numbers 1 - 5 indicate sequential order of fixation duration estimates following the luminance change. The estimates of the fixation duration sequences estimate the mean modulation for that fixation duration from the baseline (Intercept). We consider a reliable difference to be observed when ($|t| > 2$).

Distributions for the first fixation duration following the change in luminance are plotted for the two conditions (UP and DOWN) (Figure 13). Visual inspection of the empirical distributions reveals some qualitative differences between the two conditions. In the UP condition there appears to be little difference between the proportion of short fixation durations (i.e. approx. < 200ms) between the baseline and upward shift distribution. The early shift is certainly much less pronounced than in UP shift distributions observed in Chapter 2 (see upper plot Figure 9). Further inspection of the distribution reveals that there is a specific decrease in the proportion of long fixation durations starting at approximately 300 ms. It appears that the source of the decrease in observed mean fixation durations is likely attributable specifically to a decrease in long fixation durations.

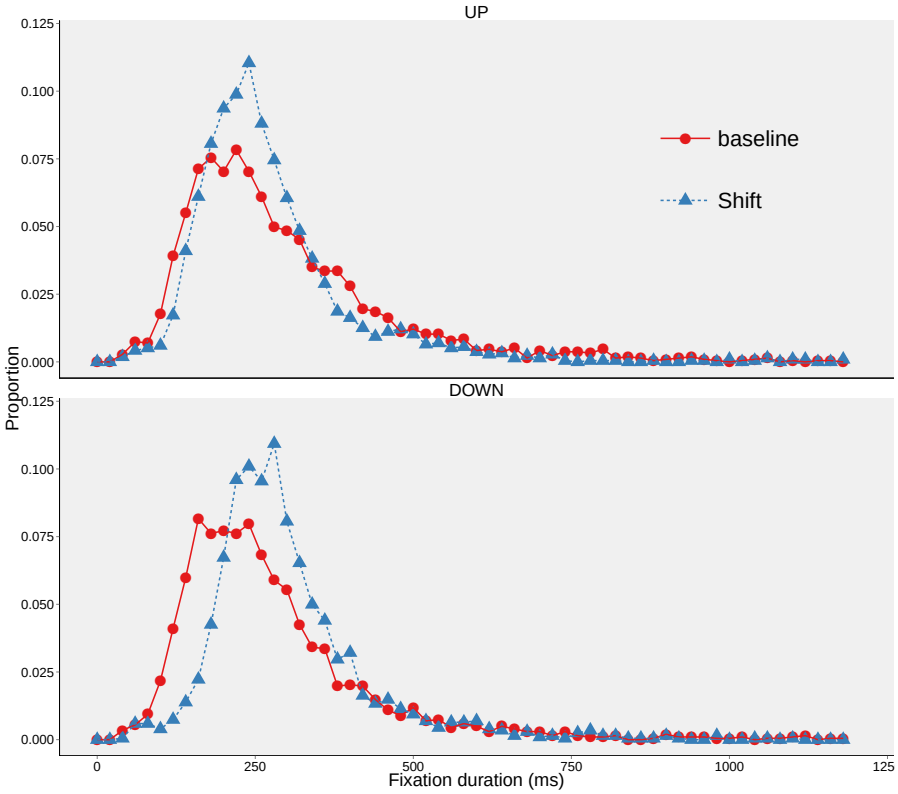

Figure 13: Mean fixation durations Experiment 4. Each point represents a fixation duration for particular fixation in a sequence. On the x-axis 0 represents the fixation immediately preceding a change in luminance. 1 - 5 represent the sequential number of the fixation that follows this change.

The distribution in the downward luminance change condition shows an early shift such that there is a tendency to observer overall fewer short fixation durations. The proportion of longer fixation durations (i.e. > 450) is comparable to the proportions observed in the baseline condition. Therefore, based on these observations it appears that the increase in mean fixation duration in the DOWN condition is associated with a decrease in the proportion of short fixation durations with no specific increase in the proportion of long fixation durations.

3.2.4 *Discussion*

We analyzed the adaptation of fixation durations to changes in scene luminance levels as a function of the ordinal fixation number relative to a change in fixation. The original aim of this study was to observe the influence that scene quality changes have on fixation durations for fixations beyond the one immediately following a change in luminance. The hypothesis that fixation durations adapt gradually to stimulus changes has been given a formal characterization in the ICAT model (Trukenbrod & Engbert, 2014). Interest in this question was generated by the interpretation of the results observed in Chapter 2 that the early increase in fixation durations is due to a surprise effect. It was reasoned that fixation durations beyond the first post-change fixation would be relatively less influenced by surprise and we aimed to quantify what effects would be present. However, in this experiment an unexpected result was observed. Fixation durations in the DOWN condition more or less followed the stereotypical pattern. Immediately following a decrease in scene quality fixation durations increased in duration. This result is consistent with what has already been reported in the present thesis (see Chapter 2) as well as empirical reports by other authors (Henderson et al., 2013; Glaholt et al., 2013). However, the observations in the UP condition represent a unexpected finding. It was predicted that due to the surprise effect that has been previously described, that we would observe a small increase in fixation durations immediately following the change. In contrast to this prediction, fixation durations were actually lower than

the baseline. Although unexpected, and not part of a formal hypothesis for this experiment, these results are consistent with what has been termed in Chapter 2 a form of symmetrical control. That is, when scene quality is decreased (i.e. luminance decrease) fixation durations adapt by increasing in duration and when scene quality is increased (i.e. luminance increase) fixation durations adapt by decreasing in duration. However, before making such a tentative conclusion it is worthwhile comparing this task with that of the tasks in Chapter 2. This is an important comparison to make because, in principle, the task are very similar but have led to very different patterns of results.

There are two primary differences in this design. The first, and most obvious, is the fact that when the luminance is increased it remains for a total of 10 fixations at the increased level of luminance. The second is a more subtle and important difference. In this experiment participants are always aware that the upcoming change will be in a specific direction. That is, when the scene is at 60% luminance the participant is always aware that the next upcoming change will be an enhancement and that this will be achieved by increasing the luminance by a fixed amount. This is in contrast to the experiments conducted in Chapter 2 where it was the case that when the stimulus was presented in baseline luminance (i.e. 60%) that the next luminance shift would be randomly chosen to be either in the upward or downward direction.

An article that was published concurrently with the analysis of these results helped to shape the interpretation of the mechanisms underlying the behaviour observed in this experiment. Henderson et al. (2014) ran a study with a similar design to that which was used in Chapter 2. Their research question was similar in that they aimed to quantify the effect that enhancements and degradations have on the control of eye-movements while viewing naturalistic scenes. In their study, they manipulated processing difficulty by modifying their scenes using a low-pass filter. A low-pass filter is method of image processing whereby the high-spatial frequency content that is above a certain cutoff is attenuated (Gonzalez, Woods & Eddins, 2009). The

high-spatial frequency content of an image is responsible for carrying information regarding the fine grained details such as lines and edges and removing these details makes it difficult to extract features composed of these details from the image. Their study utilized a block design to separately test the influence of taking away (i.e. degradation) or adding back (i.e. enhancement) high-spatial frequency content from a scene. In the degradation block, participants viewed a scene for a number of fixations and on a critical fixation high-spatial frequency content was removed from the scene. Their results in this condition was as would be expected from other reports in the literature in that fixation durations increased as the filtering became stronger (Henderson et al., 2013; Walshe & Nuthmann, 2014; Glaholt et al., 2013). However, in the enhancement condition they reported that by improving processing by adding missing high-spatial frequency information back to the image that fixation durations became shorter.

From one perspective, Henderson et al. (2014)'s results, along with the results reported in the present chapter, provide convergent evidence regarding the link between stimulus enhancements and reduced fixation durations. However, a difference between Henderson et al. (2014)'s task and the tasks in Chapter 2 lead to a hypothesis about why facilitation was observed in their study, but not ours. In their design, the enhancement and degradation conditions were blocked in such a way that stimulus enhancements and stimulus degradations were presented in separate blocks. In the enhancement blocks, a stimulus was presented at the lowest level of quality. On critical fixations, spatial frequencies were added back to the stimulus at various levels, including a condition in which the unfiltered stimulus was presented (i.e. full quality enhancement). Due to the blocked characteristics of the design the participant was always aware that the upcoming change would be a stimulus enhancement; in other words, an upcoming enhancement was predictable. It is worth noting that while it was predictable that an enhancement was upcoming, the participant was not aware of the specific *level* of enhancement. For instance, they did not know by what magnitude the stimulus would

be enhanced. Therefore, this feature of their design closely parallels a feature of the design in the present study. In both cases the luminance change is predictable. Furthermore, both studies contrast in this regard with the design of the experiments reported in Chapter 2. Based on this observation, it is speculated that the predictability of the stimulus change influences how eye-movement controls react to the stimuli. For instance, when an upcoming change is predictable we hypothesize that this should result in a decrease in the magnitude of surprise that is attached to the stimulus change. By removing surprise, facilitation that arises due to processing of a scene presented in higher quality yields shorter than average fixation durations. The distributions presented in Figure 13 provide some support for this hypothesis. In the UP condition, where we suggest that surprise is less active due to predictability we see that the early shift in the distribution is minimized in comparison to the UP distribution presented in Chapter 2, Figure 9. Unfortunately, Henderson et al. (2014) did not report distributions so a comparison to the present experiments on these grounds cannot be made.

In the experiments described in the following sections of this chapter I aim to test the hypothesis that the mechanisms responsible for reduced fixation durations under enhancement conditions is a result of stimulus predictability.

3.3 EXPERIMENT 5 - TESTING THE PREDICTABILITY HYPOTHESIS: DO PREDICTABLE LUMINANCE ENHANCEMENTS LEAD TO A DECREASE IN FIXATION DURATIONS?

3.3.1 *Introduction*

The results observed in Section 3.2 led us to the hypothesis that the eye-movement control mechanisms in scene perception are sensitive to stimulus predictability (see Section 3.2.4). Specifically, it is hypothesized that a lack of stimulus predictability was the explanation for why a stimulus enhancement lead to fixation duration increases in Chapter 2 but lead to decreases reported in Section 3.2. Convergent evidence for this hypothesis was also supplied by a concurrent study conducted by Henderson et al. (2014) in which a fixation contingent scene quality paradigm was used to vary the spatial frequencies present in an image. We argued in Section 3.2.4 that their results can also be accounted for in terms of the predictability hypothesis.

The purpose of the present Experiment is to conduct a study that is capable of testing the hypothesis outlined in that predictability is a mediating factor in determining the influence of surprise on fixation durations (see Section ch:facilsearch:section:sequence:section:discussion). This hypothesis was tested by returning to a design previously used in Chapter 2. We speculated that by making the change in luminance predictable, the effect of surprise would be minimized and fixation duration decreases would be observed. The specific method by which the change was made to be predictable was to make changes in only a single direction, namely an increase in luminance. Therefore, whenever a scene was presented in a baseline level of luminance the participant was always aware that the next change in luminance would be an increase, hence predictable. This design is very similar to that utilized in Chapter 2 with the primary difference being that the downward luminance changes have been completely removed. This design also has the added benefit in that it is conceptually very similar to the design used in Henderson et al. (2014)'s study.

3.3.2 *Methods, Equipment and Materials*

The equipment used in this study is identical to that used in the experiments reported in Chapter 2. Several small, but important, changes were made to the procedure, stimulus materials and design. A total of 12 participants were tested. The baseline luminance was set to 60% of the original scene luminance. In total, there were three luminance change conditions. Each condition was presented twice per trial, hence there were six manipulations per trial. The luminance shifts that were applied were, 60% $\rightarrow$ 60% (baseline), 60% $\rightarrow$ 80% (UP1), and 60% $\rightarrow$ 100% (UP2). All other aspects, this experiment paralleled details that are described in Section 2.2.1.

3.3.3 *Results & Discussion*

Results

The approach to analyzing the data was simple, a linear mixed-effects regression was used to statistically validate the empirical differences in mean fixation duration between the conditions. The model included random intercepts and random slopes that varied by participant. The empirical means showed mean fixation durations of 346, 386 and 378 ms in the Baseline, UP1 and UP2 increase conditions, respectively. Statistical validation revealed that the mean difference between the Baseline and UP1 condition was marginally significant, $M_{UP1-Baseline} = 38.89$, $SE = 23.03$, $t = 1.69$, and the mean difference between the baseline and UP2 condition was statistically significant, $M_{UP2-Baseline} = 31.75$, $SE = 16.13$ $t = 1.97$. The mean differences are plotted in Figure 14.

Table 7 shows the by participant as well as overall means across the three conditions. The pattern of individual responses in the 3 condition provides no indication of a trend in the data towards shorter fixation durations in any of the conditions relative to the baseline.

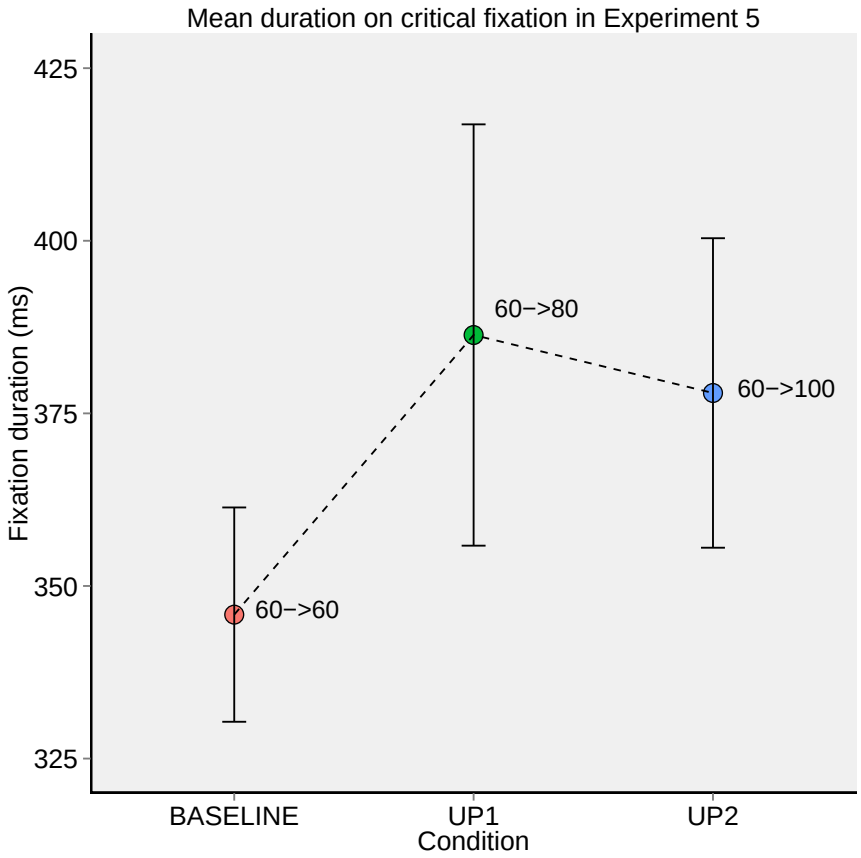

Figure 14: Mean fixation durations in Experiment 5. Baseline represents the condition in which fixation durations are measured at a baseline level of luminance (60% original scene luminance). The UP1 and UP2 conditions plot the mean fixation durations in the two luminance enhancement conditions.

Discussion

The lack of any decrease in fixation durations following an increase in luminance indicates that these specific experimental parameters are not sufficient to lead to facilitation. Given the pattern of results observed in this section, the choice was made to test a slightly different, but related hypothesis. As has been mentioned, Henderson et al. (2014) observed facilitatory effects using a experimental design that is very similar to the one presented in this section. This study was also an influential factor in leading us to the hypothesis that predictability may play an explanatory role in the results so far reported in the thesis. Conceptually, the design used in this section is a

Subject	Baseline	UP1	UP2
1	312.29	327.92	315.26
2	294.26	269.77	300.42
3	405.69	389.39	422.25
4	475.44	546.30	529.38
5	337.82	382.00	415.19
6	308.99	300.49	304.80
7	367.83	634.74	514.07
8	345.85	390.77	369.98
9	280.41	317.33	331.77
10	368.76	329.61	332.41
11	317.32	411.98	336.03
12	335.57	335.92	363.92
Mean	345.85	386.35	377.96

Table 7: Individual fixation durations and means: The baseline condition is presented at a luminance level of 60% original scene luminance. In the UP80 condition, the scene is presented at 80% original scene luminance. The original scene is presented in the UP100% condition.

variation of their Experiment 2. The primary difference is that an enhancement was used that involved increasing the stimulus luminance while they used an enhancement that involved reapplying high spatial frequencies that had been filtered from the image. In a next step, the design was modified to more closely match the design used by (Henderson et al., 2014) (i.e. spatial frequency filtered stimuli). Such a conceptual replication of their results would represent a first step in validating the assumptions underlying the hypothesis that stimulus predictability is the main driving factor underlying the facilitation that was observed in Section 3.2 and Henderson et al. (2014).

3.4 EXPERIMENT 6 - TESTING THE PREDICTABILITY HYPOTHESIS: REPLICATING HENDERSON ET AL. (2014)

3.4.1 *Introduction*

The last several sections have addressed the question of whether a decrease in fixation durations can be observed following an enhancement of stimulus quality. These questions were driven by empirical observations made from two sources. In section 3.2 facilitation was observed when luminance was shifted and then remained at the new luminance level for 10 fixations. In a study conducted by Henderson et al. (2014), facilitation was observed in the case in which a baseline image was defined by removing selected spatial frequencies from a colour scene. These two observations led to the hypothesis that the underlying mechanism governing the control of fixation durations within these experiments is sensitive to the predictability of the direction of the luminance change (i.e. increase vs decrease). This hypothesis was directly tested using luminance based stimuli in the previous section but no evidence to support the the predictability hypothesis was found.

Here, we provide a closer replication of Henderson et al. (2014) which involves using spatial frequency filtered stimuli. Given that this study provides the bedrock underlying the predictability hypothesis, it is therefore pertinent to assess the robustness of the finding via a conceptual replication. Although it is not possible to perfectly replicate the stimuli used in their original study as the filter type and cutoff frequencies used were not reported, the stimuli used in our study are comparable. The stimuli in their baseline condition are reported to contain spatial frequencies of no greater than 0.6 cycles per degree. In the baseline condition of our study, the cutoff used in the Butterworth filter was defined at 0.7 cycles per degree. Henderson et al. (2014) reported a reduction in fixation durations for the case where a baseline image is swapped for an unfiltered image. We predict that we should observe the same reduction for our baseline vs unfiltered condition.

3.4.2 *Methods, Equipment and Materials*

The equipment used in this study is identical to that used in the experiments reported in Chapter 2. The procedure and design were identical in all but several respects. A total of 13 participants were tested. A filtering process described in the Stimuli section was applied to grayscale versions of the images used in the previous experiments. Grayscale stimuli were used to eliminate any confounding influence of colour on fixation durations. Therefore, in relation to the methods used in 3.3 this study simply replaced the luminance images with grayscale spatial frequency images. All other aspects were identical.

3.4.3 *Stimuli*

A Butterworth filter was used to selectively attenuate spatial frequencies present in grayscale images. The frequency response of the Butterworth filter is given by $G^2(\omega) = \frac{{G_0}^2}{1+(\frac{\omega}{\omega_c})^{2n}}$ where n is the order of the filter, ω_c is the cutoff frequency and G_0 is the gain at zero frequency (Gonzalez et al., 2009). Stimuli were created by applying the filter to images that had been transformed into frequency space via a Fourier transformation. The Butterworth filter is a particularly useful filter in the present circumstances in that the frequency response of the filter is flat for spatial frequencies less than the cutoff but gradually declines to 0 for spatial frequencies above the cutoff. Such a gradual decline avoids the *ringing* effects present in the ideal filter, but results in a strong attenuation of signals beyond the cutoff (Gonzalez et al., 2009). A second order filter was used in the present experiment. The filtering was accomplished by a custom script that batch processed the original scene images to produce the filtered versions. The baseline condition was created by applying the filter with a cutoff frequency of 0.7 cpd. The first enhancement condition was created by applying a filter with a cutoff of 1.5 cpd. The second enhancement was the fully clarified scene.

3.4.4 *Results*

In the Baseline condition (0.7 cpd → 0.7 cpd) the mean fixation duration observed was 363 ms. In the UP1 condition (0.7 cpd → 1.5 cpd) the mean fixation duration was 437 ms and in the UP2 condition (0.7 cpd → Unfiltered) the mean fixation duration was 405 ms (see Figure 15). The individual mean fixation durations are provided in Table 8. A linear mixed-effects regression was used to statistically validate the difference in mean responses observed in the enhancement conditions relative to the baseline conditions. Random intercepts and random slopes that varied by subject were included. Other details regarding the regression modelling can be found in 2.2.1. The increase in the first enhancement condition was marginally significant $M_{UP1-Baseline} = 74.11$, $SE = 43.59$ $t = 1.70$. The fixation duration increase in the second enhancement condition was not statistically significant $M_{UP2-Baseline} = 41.25$, $SE = 37.45$ $t = 1.10$.

Subject	Baseline	UP1	UP2
1	379.14	462.75	362.72
2	317.99	292.33	272.54
3	358.55	807.42	632.88
4	291.32	394.00	333.50
5	534.99	515.27	463.70
6	399.41	772.71	730.07
7	381.31	343.12	305.22
8	336.66	297.73	268.84
9	289.73	294.26	268.12
10	311.82	296.54	301.90
11	289.91	331.01	413.90
12	529.19	537.36	483.72
13	297.34	339.47	422.26
Mean	362.87	437.23	404.57

Table 8: Individual fixation durations and means: The baseline condition is presented with spatial frequencies above 0.7 cpd strongly attenuated. In the UP1 condition, a clarified image is presented such that only spatial frequencies above 1.5 cpd are strongly attenuated. In the UP2 condition, the unfiltered image is presented.

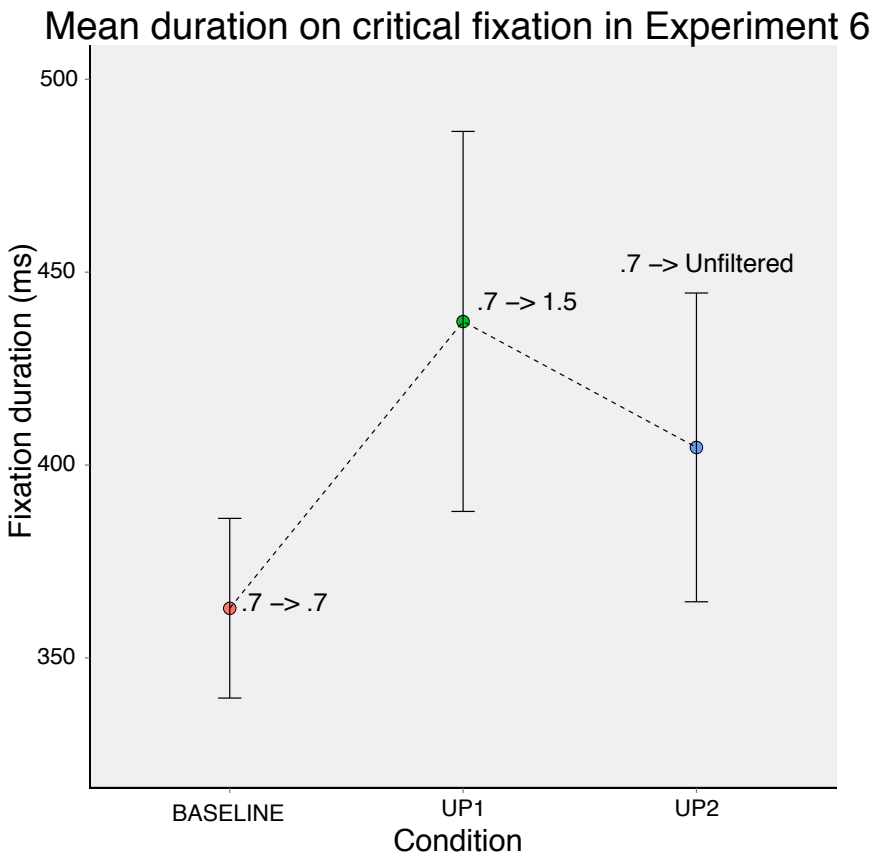

Figure 15: Mean fixation durations in Experiment 6. Baseline represents the condition in which fixation durations are measured at a baseline level of spatial frequency filtering (0.7 cpd). The UP1 and UP2 conditions plot the mean fixation durations in the two spatial frequency enhancement conditions.

3.4.5 *Discussion*

Given the lack of any observed decrease in fixation durations between the baseline and fully clarified scene we do not replicate the results reported in Henderson et al. (2014). Given that the effect reported in their paper was quite strong (90 ms) we anticipated having no difficulty replicating the study using stimuli and a design which approximated their own. Based on our failure to replicate, we believe that this effect is more subtle and less generalizable than we had previously anticipated. Although a decrease in fixation durations with an increase in stimulus quality is clear from the data reported in Henderson et al. (2014) this decrease may be strongly tied to the particular characteristics of their stimuli and design that were not replicated. The replica-

tion that we conducted can be seen as a stripped down version of the original study, so it is therefore pertinent to compare what features the original Henderson et al. (2014) study contained that were not included in our replication. It may be the case that these factors explain the difference in the results observed in our replication and the original study. For the immediate next step along this avenue the most natural course would be to not only replicate this study in a conceptual manner, as was done in this section, but to replicate it in as close a format as possible to the original. In the following, several salient differences between the original and our replication are highlighted to guide future work aiming to replicate this theoretically important, but experimentally unstable, result.

FILTERING From the reporting in the original paper it is difficult to determine exactly what type of filtering had been carried out. For instance, the specific filter that was used was not reported. They reported their stimuli had been filtered to 0.6 cpd, but it is not clear how this was accomplished. Spatial frequency filters are not all equal in their frequency response profile. For instance, the rate at which spatial frequencies attenuation drops off beyond the cutoff may vary between filters. However, it is unlikely that such a difference between the filter alone would explain the difference. By visually inspecting sample images presented in their paper it is apparent that the filter used to construct the baseline stimuli is roughly comparable with our stimuli. While it is unlikely that the difference in filter directly accounts for the difference in fixation durations this remains a possibility. A future replication could aim to find out the exact filter properties used and adjust the stimuli used in the replication accordingly. Alternatively, the stimuli used in the original study could be requested and the replication could be conducted on the identical stimuli.

COLOUR In the original study, spatial frequency filtering was applied to colour images whereas our replication used images reduced to grayscale channels only. It may be the case that the mechanism

responsible for decreased fixation durations following the addition of spatial frequencies is mediated by distinct neural pathways that process colour information (Gegenfurtner, 2003; Shevell, 2008). Future work could address the influence of colour conducting a further experiment using the same experimental design but using full colour stimuli.

NUMBER OF ENHANCEMENTS In our study we opted for only two spatial frequency enhancement conditions. In the original study enhancement was observed only between the baseline level and the unfiltered images. Since our goal was to replicate this basic enhancement effect we reduced the number of enhancement steps used in their study from four to two. If what Henderson et al. (2014) observed is a true enhancement effect it is difficult to understand why increasing the number of enhancement conditions would explain the difference in results. Nevertheless, in a future replication the number of conditions and the relative difference between spatial frequencies across conditions should be consistent with the original study.

3.5 SUMMARY AND DISCUSSION OF MAIN EXPERIMENTAL FINDINGS.

Part II of the thesis represents a detailed examination of visual information processing mechanisms that are engaged to govern the control of eye-movements during naturalistic scene perception. This examination focuses in particular on the behaviours of the eye-movement control system that operate on the time-scale of a single fixation. That is, when a fixation is made on a region of a scene, how does the duration of the fixation adapt to reflect the characteristics of the stimulus content that is encountered. This general background led to a simple testable hypothesis: Do transient changes in scene luminance that are made during a saccadic eye-movement, and which last for the duration of the subsequent fixation, lead to both increases and decreases in fixation durations? An eye-movement control mechanism that is sensitive to, and capable of immediately exploiting, increases in stimulus quality would be predicted to lower its fixation durations following an increase in scene luminance. Surprisingly, this link between increases in stimulus quality and the resulting impact on fixation durations had not been directly tested within the context of scene perception tasks when the project was formulated. The recent convergent (and divergent) evidence that has been reported on this question (see Walshe & Nuthmann, 2014; Henderson et al., 2014; Glaholt et al., 2013) provides an indication of the importance and relevance of the question to ongoing debates in the psychology of eye-movement control. In the section that follows, the key implications from the pattern of results observed in the studies will be summarised and discussed. The most salient aspects of this pattern are extracted to inform the development of model simulations that are reported in next chapter (see Chapter 4). This section also includes a series of general recommendations for how future work could improve upon the overall methodology.

In Chapter 2 we observed fixation duration increases under conditions in which stimulus information (luminance) was both increased and decreased. The fixation duration increase that followed a lumi-

nance decrease replicated a pattern previous reported in Henderson et al. (2013). The small increase in fixation durations that was observed in the luminance increase condition was contrary to expectations. Increasing the available information via a luminance increase was predicted to result in no change or a small decrease but not an increase in fixation durations. Work conducted concurrently by Glaholt et al. (2013) provided several clues as to why such a pattern had been observed in our data. Using a gaze-contingent scene quality approach, fixation durations were increased following several types of quality changes. Increases in fixation duration were also found when missing colour channels were reapplied to the stimuli [1]. This increase in fixation durations for colour stimuli parallel the results that have were reported in Walshe & Nuthmann (2014) for luminance stimuli. The authors constructed an argument based on the effects that their manipulations had on the distribution of fixation durations to suggest a dual-process account for their data. The first influence, which is fast acting and influences even the shortest fixation durations, is referred to as a surprise effect. Such a surprise effect results from a detected mismatch between the pre- and post- saccadic stimulus content. The second influence is referred to as processing related. This influence is hypothesized to occur relatively late within a fixation due to the time required for stimulus identification to be completed. Therefore, processing related influences are hypothesized to influence only relatively long fixation durations. Short fixation durations are unaffected by the processing related influences as the stimulus is not processed quickly enough to exert an influence on these short duration fixations. Glaholt et al. (2013)'s results were consistent with such an account in that manipulations in which stimulus processing was made more difficult resulted in a specific influence on long fixation durations. However, manipulations in which no such processing difficulty increase was expected did not yield such specific increases to the long fixation durations. However, a general increase in fixation durations was observed even for very short fixations. The analysis of the distributions from Experiments 1 and 2 are consistent with such a dual-process

1 For a more detailed discussion of this work, see 2.1

model. The pattern of fixation duration increases was consistent with such an interpretation. In the DOWN condition, where the decrease in luminance was anticipated to result in processing difficulties, we observe a specific influence on long fixation durations - as was observed in Glaholt et al. (2013)'s studies. Importantly, in the UP condition, where no such difficulties are anticipated, no specific influence on the tail of the distributions was observed. Rather, the increase in fixation durations was consistent with a fast acting effect of surprise that shifts the mode of the distribution and therefore acts on even the shortest fixation durations. The interpretation of the results from Glaholt et al. (2013) and Walshe & Nuthmann (2014) lead directly to the claim that the mechanisms underlying fixation duration control in scenes are influenced by an effect that we have previously referred to as *surprise*.

If surprise is a general characteristic of the response that the eye-movement control system has to visual transients, are there contexts in which the enhancement of stimulus content may overcome the degree of surprise generated by the change? The results on this account are mixed. Recently, Henderson et al. (2014) published a gaze-contingent scene quality manipulation study in which the spatial frequency of the stimuli were manipulated. Stimulus enhancement was accomplished by replacing spatial frequencies that were filtered from a scene during a selected saccadic eye-movement. The surprise hypothesis would predict that the eye-movement control mechanisms would be inhibited by the visual change elicited by changing the scene. However, the authors reported a statistically significant reduction in fixation durations from the condition in which the filtered scene was replaced with a completely unfiltered scene. This means that if such a surprise effect was present, as would be expected, the facilitation in processing resulting from the addition of spatial frequencies was large enough to overcome the inhibitory surprise. Caution must be applied when interpreting these results. In Section 3.4, we attempted to replicate these results under similar conditions but were not able to reproduce the reported decrease in fixation durations. In that section, we suggest several factors that may contribute

to the difference in results. We believe that at this time it would be ill-advised to make strong inferences based on Henderson et al. (2014)'s results. Nevertheless, with caution, these results may be suggestive that under certain particular circumstances the eye-movement control system may decrease fixations in response to an increase in stimulus quality.

One study reported within the present chapter did show a data pattern consistent with a decrease in fixation durations following an increase in stimulus quality. Experiment 4 reported in Section 3.2 found that fixation durations decreased on the first fixation following a gaze-contingent increase in stimulus luminance. In this task, the increase in scene luminance was left intact for 10 fixations, unlike in previous studies. In addition, the direction of the upcoming luminance shift was easily determined by the participant. We hypothesized that due to the predictable nature of the stimulus change, the magnitude of the surprise effect is reduced yielding shorter fixation durations in the case that a increase of stimulus quality is obtained. This hypothesis also gained strength from the observation that in Henderson et al. (2014)'s study, the participants were aware that all changes would *add* spatial frequencies. We predicted that if we modified the design of Experiment 1 and Experiment 2 to make the luminance change predictable that the pattern of fixation durations would be consistent with Experiment 4 and Henderson et al. (2014)'s Experiment 2. As is reported in Section 3.3, contrary to our prediction, a reduction in fixation durations was not observed. As was previously mentioned, to make our results more comparable to those reported by Henderson et al. (2014), we also replicated this approach using spatial frequency filtered stimuli. However, we do not report a replication of the results reported in their paper.

In summary, there is convergent evidence that surprise plays a role in the response that the eye-movement control system has to gaze-contingent changes in scene quality. This evidence comes from studies reported within this thesis chapter, as well as a number of recently reported empirical results (Walshe & Nuthmann, 2014; Glaholt et al., 2013; Henderson et al., 2014). There are early indications that the

improvement in stimulus quality may overcome the effect of surprise under certain experimental conditions and lead to reduced fixation durations. However, as has been discussed, these results must be regarded with caution as the experimental effects have not generalised well beyond the specific experimental conditions in which they were observed.

Part III

COMPUTATIONAL MODELLING OF FIXATION DURATIONS IN SCENE VIEWING

In the final part of the thesis a theoretical framework is introduced to provide an explanation for empirical data presented in other section of this thesis. The theory is developed in the form of a computational model and is therefore capable of making explicit testable predictions. Among the many strengths that a computational theory has over a verbal theory is the ability to simulate data and describe in exact detail what the model predicts. The model is applied to data reported within this thesis and the resulting behaviour is analyzed in detail.

4

THE UCM MODEL

Well, sir, it's this rug I had. It really tied the room together.
— Jeff "the Dude" Lebowski

4.1 INTRODUCTION

The study of the relationship between visual processing and eye-movements presents unique opportunities to decipher the underlying mechanisms that mediate this relationship. As information is provided to the visual system, mechanisms are engaged that result in systematic changes in behaviour. By observing these patterns it is possible to make inferences about the information processing operations that these behaviours rely on.

Directly controlled fixation durations

A topic of debate is the degree to which eye-movements are directly-controlled by the stimulus content upon which gaze is currently directed. With regards to the timing of eye-movement decisions, research aims to determine the degree to which processing of content currently being assessed within a fixation may intervene in influencing the timing of the next saccadic decision. Accounts of eye-movement control that adopt strong direct control assumptions suggest that stimulus processing occurs rapidly enough to have such an influence and that the timing of eye-movement decisions is governed by this processing. Empirical evidence has provided a great deal of support for the direct control hypothesis. Research on reading behaviour has shown that fixation durations increase when a fixation lands on a word with low predictability or low frequency (Kliegl, Nuthmann & Engbert, 2006) or when the text is presented in visu-

ally degraded format (Glaholt et al., 2014). In scene perception, direct control effects on fixation durations have been observed under various conditions such as when a fixation encounters a luminance degraded scene (Henderson et al., 2013; Walshe & Nuthmann, 2014) or when spatial frequency content has been removed from the stimuli (Glaholt et al., 2013; Henderson et al., 2014). Such effects have also been observed for fixations that land on a task relevant vs irrelevant stimuli (Glaholt & Reingold, 2012) as well as fixations in which a task irrelevant distractor stimulus is presented (Pannasch et al., 2011). Overall, compelling evidence exists to support the claim that some proportion of fixations are directly controlled by stimulus content.

Despite such evidence for direct control, research has demonstrated that not all fixations show such an effect. Fixation durations that do not appear to be influenced by currently fixated stimulus properties are often referred to as being indirectly controlled. Indirect control postulates that the duration of an individual fixation is determined by factors that are external to the local stimulus content encountered within a fixation. Instead, fixations may be executed at random intervals with a mean rate set by global factors external to the fixated stimulus (for a summary see Reingold et al., 2012). In a study of direction-coded search, Hooge & Erkelens (1998) argued that performance failures could be explained by an inability for participants to match fixation durations to the processing demands of the fixated stimulus. Such a result violates the assumptions of a pure direct control eye-guidance mechanism. Henderson & Pierce (2008) used a technique known as the stimulus onset delay paradigm (SOD) (also see Shioiri, 1993) to investigate the control of fixation durations during naturalistic scene perception. In this paradigm participants actively explored the scene and on a selected saccade the entire scene was removed from view. After a variable delay within the subsequent fixation, the scene was replaced to view. Consistent with a direct control account, they found that one population of fixation durations increased to match the length of the delay. However, they found that another population of fixation durations was unaffected by the length of the SOD. On this basis, the authors argued for a mixed-control ac-

count with one population being indirectly controlled and the other being directly-controlled.

CHALLENGES FOR DIRECT CONTROL A serious challenge that direct control theories must face is how such effects are at all possible given the limited processing time that is available for stimulus processing to influence eye-movements (for a review see Reichle & Reingold, 2013). In a typical naturalistic viewing task, fixation durations are approximately 300 ms with a wide variability around this mean; however, direct control effects have substantially less time to exert their influence. This becomes apparent when the timing of specific neural events related to stimulus processing are taken into account (Reichle & Reingold, 2013). It is known that approximately 60 ms is required to transmit visual signals represented on the retina to regions of the brain responsible for processing these signals. Furthermore, an additional 90 ms is required to encode these signals and process them into a format in which they may be interpreted. The latency of the movement signal to pass from the brain to eye is thought to be approximately 20 ms (Becker, 1991). Given these timing constraints, any direct control effect of stimulus content on the timing of the resulting saccade must logically occur between the time needed for the stimulus to reach brain areas responsible for saccadic decisions and onset of transmission of the oculomotor movement signals. Any theoretical approach to understanding the underlying mechanisms must give consideration to such timing constraints.

Eye-movement control models in high-level tasks

Eye-movement control theories have been most actively developed within the study of reading behaviours (Engbert et al., 2005; Reichle et al., 1998; Legge, Klitz & Tjan, 1997; McDonald, Carpenter & Shillcock, 2005). Reading presents fertile ground for testing such theories as movement trajectories follow highly stereotyped patterns and the stimulus varies in predictable and easily measurable ways. However, recent models of the temporal properties of eye-movement decisions

have been successfully extended beyond the reading domain (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014; Reichle et al., 2012).

THE CRISP MODEL The CRISP model was developed as a model of eye-movement control during naturalistic scene perception, but has also been applied to modelling fixation durations in reading tasks (Nuthmann & Henderson, 2012). Saccade programs are initiated at a rate determined by an autonomous timing process (Engbert et al., 2002, 2005). In the CRISP model, the timer was implemented as a stochastic random walk towards a threshold. Once the threshold value is reached, a saccade program is initiated that may then, eventually, terminate the fixation. The CRISP model may be referred to as an example of a mixed-control mechanism (c.f. direct control) due to the manner in which within-fixation stimulus influences impact the timing signal. The direct control component of the CRISP model arises from two sources. The first is due to inhibitory influences that may interfere with the rate at which the timing signal approaches threshold. For example, if a complex or difficult to process stimulus is fixated, the timer rate may be inhibited. The second direct control influence in the CRISP model arises when stimulus processing elicits a cancellation of a saccade that is within the initial stages of preparation. On the other hand, indirect control occurs in the CRISP model when saccade programming is initiated prior to the onset of the current fixation and when no processing related cancellation occurs. In these cases, since a saccade program has been initiated prior to the onset of the current fixation, modification of the timer rate will not have any impact on the current fixation duration. Nuthmann et al. (2010, see Figure 4) showed that advanced preparation saccades tend to be those that produce short fixation durations. Therefore, given this fact the CRISP model implicitly predicts that short fixations tend to avoid the influence of direct control.

ICAT ICAT is a model of eye-movement control in high-level tasks that utilizes similar principles to the CRISP model. Like the CRISP mode, saccade programs are initiated according to a random timing

process that may be subject to processing related interference. Once the timer reaches threshold, a labile program is initiated. However, unlike the CRISP model, ICAT implements the labile and nonlabile stages of saccade programming as independent random walk processes. In ICAT, the duration required for saccade programming is equal to the sum of the durations of each independent random walk. A novel aspect of ICAT is that it permits interference of saccade programs at the level of the labile stage of programming. When foveal processing difficulty increases, the timer rate of the labile stage of programming is decreased. This is in contrast to the CRISP model which does not permit the modification of labile durations in response to task demands.

ICAT also makes more elaborate assumptions about the relationship between stimulus processing and the duration of saccade timing and saccade programming processes. This elaboration is accomplished by introduction of local[1] vs global control principles. Changes to the mean rate of the saccade timer that occur at a global level are not tied to the immediate processing context (i.e. foveated content). Rather, these global effects are sensitive to properties of the stimulus encountered on previous fixations and reflect general aspects of the difficulty of the task. Local control is the mechanism implemented in the ICAT model to control changes in the rate at which the saccade timer and labile stage of programming adapts to currently foveated content[2]. Thus, ICAT's local control mechanisms play a similar role to the direct control mechanisms implemented in the CRISP model. ICAT's local control process is responsible for adaptive changes to timer rate that occur on the level of a single fixation.

1 Within ICAT this is technically referred to as Local-II. Local-I is a term that refers to variability in mean duration that results from the stochasticity of the random walk. Our focus is on local-II as we are interested in changes that occur to the rate of the timing process

2 It should also be noted that global and local control interact. Global control mechanisms scale local inhibition depending on the anticipated processing difficulty.

Asymmetrical control of fixation durations

Research from reading and visual search tasks has led to the suggestion that there is an asymmetry in how eye-movement control mechanisms respond to increases and decreases in foveal processing difficulty. In reading, Kennison & Clifton (1995) investigated the impact of word frequency on two adjacent words embedded in single sentences. High and low word frequency adjectives were followed by high and low word frequency nouns. Parafoveal preview of the noun was prevented by using the invisible boundary technique. When readers first fixated a high-frequency adjective, fixation durations on the subsequent noun showed a word frequency effect, such that longer fixation durations were observed for low-frequency than for high-frequency nouns. In contrast, no such word frequency effect was observed when readers first fixated a low-frequency adjective. Thus, increasing processing demands (high $\rightarrow$ low) resulted in an immediate prolongation of fixation durations, whereas decreasing processing demands (low $\rightarrow$ high) showed no immediate facilitatory effect. Such an asymmetry in the temporal control of fixation durations has also been observed in visual search. Hooge et al. (2007) used a search task in which participants were required to find a closed ring amongst distractor Cs. The distractors in their task varied in the size of the gap, such that small gap Cs were more difficult to distinguish from the target stimulus than were large gap Cs. They found that fixations on small gap Cs that were preceded by a fixation on a large gap C showed increased durations. However, a fixation on a large gap C following a fixation on a small gap did not show a corresponding decrease in fixation durations. Walshe & Nuthmann (2014) used a task in which scene quality (i.e. luminance) was gaze-contingently modified during the saccade preceding a critical fixation (Henderson et al., 2013). They found that when the quality was increased (luminance increase) and decreased (luminance decrease) that fixation durations increased. Taken together, these results are suggestive that such an asymmetric pattern of fixation durations is observed due to the asym-

metric nature of the underlying mechanisms used to control fixation timing.

Recently, several studies have been conducted that suggest an alternative to strictly asymmetrically controlled fixation durations. Glaholt et al. (2013) showed direct control effects on fixation durations that depended on whether the image had been low- or high-pass filtered. In their main experiment, during a selected critical fixation the (gray-scale) scene was changed to a high-pass or low-pass spatial frequency filtered version.[3] Under both conditions, fixation durations increased, with both filter conditions resulting in a general shift in the mode of the distributions that occurs for even very short fixation durations. However, it was also found that low-pass filtering produced a larger effect on fixation durations than did high-pass filtering. Using distributional analysis the authors showed that the difference between the two conditions arises primary from an increase in long fixation durations (longer tail) but not from a difference in mode between the two distributions.

On the basis of these results, the authors presented a dual-process account to explain the source of these two distinct distributional effects. Specifically, Glaholt et al. (2013) suggested that rapid influences on the mode of the distribution are the result of a *surprise* effect. The surprise effect is fast acting, modifies even the shortest fixation durations, and results from a mismatch in pre- and post-saccadic stimulus content. Changes in the tail of the distribution were observed only for experimental conditions in which the change resulted in an increase in processing difficulty. On this basis, the authors argued that increases that occur specifically on the the tail of the distribution result from *processing related* influences on fixation durations. A similar account was suggested by Walshe & Nuthmann (2014) to account for the distributional effects observed in their study with luminance stimuli. Consistent with Glaholt et al. (2013)'s dual-process account, they found that while fixation durations tended to increase even when the scene was made easier to process, this increase came from a general

3 The change occurred during the saccade prior to the fixation while processing of new visual information is suppressed (Ross et al., 2001).

shift in the mode of the distribution and not from any influence on the tail.

A question that arises naturally from the dual-process account is whether it is possible for an increase in scene-quality to be substantial enough to overcome any surprise effect that results from the detected mismatch in scene features. This question was recently addressed by Henderson et al. (2014) in a saccade contingent scene-quality change study on how spatial frequency changes impact fixation durations. Participants viewed scenes which had been reduced in quality by low-pass filtering the scene. On every 6^{th} saccade the reduced quality image was replaced by one of four clarified (i.e. less strongly filtered) stimuli or the original unfiltered scene. What they found in the unfiltered scene condition was that fixation durations *decreased* (90 ms) relative to fixation durations on the baseline stimuli. In another experiment, the authors presented participants with base stimuli that consisted of fully unfiltered scenes. In this study, spatial frequencies were *removed* rather than *added*. They observed that fixation durations *increased* when a filtered stimulus was presented. Such a pattern of results is contrary to what would be predicted from a purely asymmetrical account. For this reason, the results of Henderson et al. (2014) and Walshe & Nuthmann (2014) and partially at odds. Using a similar experimental design, Henderson et al. (2014) observed symmetry while Walshe & Nuthmann (2014) observe asymmetry. However, the dual-processing account suggests a possible explanation. Specifically, if the surprise induced by the scene change was overcome by late encoding related facilitation, then a decrease in fixations may be observed. Unfortunately, Henderson et al. (2014) did not present distributional analyses that are required to confirm this prediction from the dual-process account.

A dual-process model of fixation duration control during scenes

In the following, a computational model (The UCM model) is formulated that merges the dual-process account with recent models of eye-movement control in high-level tasks. Specifically, we borrow

principles from both the CRISP model (Nuthmann et al., 2010) and from ICAT (Trukenbrod & Engbert, 2014) and supplement these with principles derived implicitly from the dual-process account. Thus, the computational model is used to make these implicit dual-process assumption explicit by simulating distributions of fixation durations and comparing them against empirical observations.

In brief, the UCM model is based on a stochastic simulation approach that uses sequences of randomly generated saccade timing signals to initiate the programming of saccadic eye movements (Engbert et al., 2002, 2005; Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). In the UCM model, both saccade timing and saccade programming are modelled as stochastic random walk processes (Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). This means that the durations required to initiate and program a saccade are partially determined by the inherent unpredictability of the random walk. The decision to move the eye from the current fixation is made once saccade programming is completed. The dual-process account predicts two methods by which the mechanisms responsible for gaze shifts may be modified on a moment-to-moment basis (i.e. surprise and encoding modulation). Thus, the dual-process assumptions enter the model at the level of saccade programming. That is, both surprise and encoding related signals can inhibit the rate at which saccade programs are completed. We also leave open the possibility for encoding related facilitation of saccade programming (Henderson et al., 2014). In what follows, we introduce model mechanisms first in an informal then in a formal manner and then use the model to simulate data reported in Walshe & Nuthmann (2014).

4.2 HIGH-LEVEL MODEL SUMMARY

4.2.1 *Rhythmic saccade timer*

Rhythmic timing is a core principle in the UCM model and is shared by other models of oculomotor control in high-level tasks (Engbert et al., 2002; Nuthmann et al., 2010; Schad & Engbert, 2012; Truken-

brod & Engbert, 2014). These models all share the basic assumption that the mechanisms underlying eye-movement decisions rely on processes that generate behaviourally relevant signals at random intervals defined by a mean rate of completion. In the UCM model, rhythmic timing plays a role in governing the signal that initiates the programming of a saccade. The rhythmic timer is conceptualized as a mechanisms that is continuously active and initiates saccade programs independently of the currently fixated content. Thus, the execution of saccade programs by the saccade timer is an indirect control influence on fixation times due to the full independence of the timer from ongoing stimulus processing. The assumption that the saccade timer operates independently of any processing demands diverges from the interference mechanism introduced by the CRISP model and ICAT. In both models, processing related interference is assumed to inhibit the rate at which the saccade timer triggered saccade programs. In contrast, the UCM model implements a fully autonomous rhythmic timer.

4.2.2 *Saccade programming*

LABILE VS NONLABILE STAGES Saccade programming is completed in multiple distinct stages of processing. Each stage of saccade programming is implemented as an independent random walk towards threshold. This multi-stage saccade programming assumption is derived from empirical investigation into saccade programming in both simple (Becker & Jürgens, 1979; Ludwig et al., 2007) and high-level tasks (Walshe & Nuthmann, 2013, 2015). Using a technique known as the double-step paradigm, these studies have shown that saccade programs that are currently underway may be updated when novel stimulus information is presented. In a typical double-step task a target is presented to an individual to which they are required to make a saccade. On some trials, after a short delay, this target moves to a new location. The individual is given instruction that in the case that a target moves, that they should always move their eyes to the new location, if possible. It is typically found that

if the target moved within approximately 80 ms of the onset of a saccade, that the response saccade will be very likely to land on the old target location. This fact was used to suggest that any new target information that occurs within 80 ms of the onset of a saccade is within what has been called the saccadic dead time. In other words, saccade programming has passed the point-of-no-return. Based on these results, Reichle et al. (1998) suggested that there is a labile and a nonlabile stage of saccade programming. The labile stage is characterized by the possibility for labile saccade programs to be subject to cancellation. However, when a saccade program has passed from the labile to nonlabile stage it may no longer be cancelled. This principle has since been used in numerous models of high-level tasks (Engbert et al., 2002, 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014).

The division of labile and nonlabile saccade programs is also adopted in the UCM model. As such, saccade programs that are within the labile stage are subject to cancellation. Saccade cancellation occurs in the UCM model when the saccade timer initiates a new labile saccade program while a previous labile saccade program is currently active. In this case a new labile program is activated that replaces the ongoing labile program.

CONCURRENT SACCADE PREPARATION It is known that saccades may be programmed in parallel, a result which may also be inferred from studies using double-step targets (Becker & Jürgens, 1979; McPeek, Skavenski & Nakayama, 2000; Caspi, Beutter & Eckstein, 2004) and neurophysiological data (Shen & Paré, 2014). Concurrency in saccade programming is another inference that is made from results using double-step stimuli. In double-step tasks, saccades that have not yet passed the point-of-no-return often show landing positions that are between the old and new target positions. It is typically found that when a saccade lands between the two target positions a corrective saccade is made to the final target position. However, concurrency is revealed in the reduction in the time required to program this corrective saccade. This inter-saccadic interval is typically well below the duration of an ordinary saccade and decreases linearly with the

amount of time that is available to processes the corrective saccade with the initial saccade (McPeek et al., 2000; Camalier et al., 2007).

In the UCM model, parallel programming occurs when the saccade timer initiates a labile saccade program while a nonlabile program is currently activated. Such parallel programming can lead to shorter fixation durations on the subsequent fixation due to the partial preparation of a saccade plan that has been engaged on the previous fixation.

4.2.3 *Dual-process saccade programming modulation*

In the UCM model, influences of direct control occur at the level of the programming of saccades and not at the level of the saccade timer. While the saccade timer operates as a fully autonomous timing process, saccade programming is subject to time dependent modulations. In effect, this implies that there are changes to the rate at which the labile and nonlabile stages accumulate towards threshold. In the model, rate modulation may occur for both the labile and nonlabile accumulators.

The dual-process account suggests that there are two distinct direct control influences on the timing of fixations. The first influence has been referred to as a surprise effect (Glaholt et al., 2013; Walshe & Nuthmann, 2014). This effect occurs when the eye lands on a location which contains visual features that strongly depart from what was expected prior to the onset of the eye-movement(Glaholt et al., 2013). Due to it's rapid onset, surprise can influence even very short fixation durations. The surprise effect is purely inhibitory and results in a slow-down in the rate of both the labile and nonlabile programs that occurs at the minimal delay in transmission of signals from retina to brain (50 ms) (Reichle & Reingold, 2013).

The dual-process account also predicts direct control effects that occur towards the later stages of stimulus processing. These encoding related effects arise when difficulties in stimulus processing are encountered. In the UCM model, encoding modulation occurs due to changes in labile and nonlabile programming rate. However, while

surprise is fast acting, encoding modulation occurs only towards the later stages of stimulus processing. Therefore, only relatively long fixation durations will be subject to this influence. There are two types of encoding modulation that are possible within the UCM model. The first occurs when saccade programming is inhibited due to difficulties encountered in stimulus processing. Such inhibition effects are motivated by numerous studies showing that fixation durations increase to match the difficulty of current scene content (Henderson et al., 2013; Walshe & Nuthmann, 2014; Hooge et al., 2007). The second type of encoding modulation occurs when saccade programming is facilitated. This can occur when the stimulus is relatively easy to process. This mechanism is motivated by recent research in scene perception which suggests that eye-movement mechanisms may operate in a bidirectional manner (Henderson et al., 2014). It is important to note that facilitation is included as speculation about how bidirectional adjustments could operate within the context of saccade programming rate adjustments. However it plays very little role in explaining the distribution of fixation durations under examination here.

4.3 FORMAL DERIVATION OF THE MODEL

Here we develop the mathematical formulation of the model architecture. The model shares many core features with other random walk approaches to modelling fixation durations during high-level tasks (Schad & Engbert, 2012; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). A unifying feature of these models is that a random walk process is implemented to account for stochastic variability at the level of saccade initiation intervals. Following previous models (Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), we implement the saccade timer as a discrete-state, continuous-time Markov process (Gillespie, 1978). First, we derive the random walk process for the case where the saccade initiation timer is the only active mechanism. Later, we extend the model derivation for the case in which multiple saccade programming mech-

anisms are activated in parallel (see also Trukenbrod & Engbert, 2014, appendix).

4.3.1 *Timer only random walk*

A random walk in state m, at time t, is given by $S_m(t)$ with initial state given by $S_0(0)$. An elementary transition occurs when the random walk changes state from S_m to an adjacent state S_n. The random walk continues until it reaches a treshold value such that $n = N$. Once the random walk reaches threshold it is reset to the initial state S_0. Elementary transitions of the random walk occur at continuously valued intervals. Therefore, a random walk that is in state m at time t will transition to the next state n at time $t + \tau$: $S_m(t) \rightarrow S_n(t + \tau)$. The duration that the random walk pauses in a given state is referred to as the waiting time. The mean waiting time τ of a timer step is related to the transition probability rate of the random walk through the equation

$$w_1 = \frac{N_{timer}}{T_{timer}}, \tag{2}$$

where w_1 governs how quickly the random walk transitions between adjacent states. The rate parameter is equal to the proportion of the total number of states N_{timer} of the timer random walk to the expected duration required for the timer to reach threshold T_{timer}.

For a discrete-state continuous-time process Markov process with constant transition probability rate with respect to τ, the probability distribution defined over values of τ is given by the exponential distribution

$$\rho(\tau) = W_n e^{-W_n \tau}. \tag{3}$$

In the simple case, in which only the random walk of the timer is active, the rate parameter of the exponential distribution W_m is given by the transition probability rate of the random walk such that

$W_m = w_1 = w_{timer}$. Specific realizations of τ may be obtained by applying the function

$$\tau = \frac{1}{W_m} \log(1 - \epsilon), \tag{4}$$

where ϵ is a pseudorandomly generated number over the interval $0 \leqslant r \leqslant 1$ (Gillespie, 1978).

The formulation provides a complete description of the development of the random walk process under the conditions that the random timer is the only active stochastic process (see **?**).

4.3.2 *Multiple Random Walks*

To fully simulate the mechanisms governing the timing of eye-movements we use a framework capable of describing stochastic accumulation at multiple independent levels. This method has previously been applied to a model of visual search and reading (Trukenbrod & Engbert, 2014) and is a generalization of the Monte Carlo approach used to simulate the discrete-state continuous-time Markov process introduced previously for the timer-only model (Gillespie, 1978). We model five independent one-step processes to account for stochastic variability in the preparation and execution of saccadic eye-movements. We also fully describe how properties of the environment are modelled by parameters of the random walk.

The composite description of the state of the model at time t is given by the vector $S_m(t) = (m_1, m_2, m_3, m_4, m_5)$[4] where $\{m_i \in \mathbb{N} | m_i \leqslant N_i\}$ and N_i is the number of steps in the i^{th} random walk. Therefore, the overall dynamic state of the random walk is determined by the state of the each of the elementary random walks. Transitions between states in the model are given by S_{nm} and is equivalent to a single elementary state change $m_i + 1$, for some $i \in 1 \ldots 5$. In effect, a change in state occurs when an elementary random walk accumulates by one step towards it's threshold value N_i. An elemen-

4 m_1 gives the state of the random timer, m_2 the labile state, m_3, the nonlabile state, m_4 the efferent transmission time, m_5 saccade execution time.

tary random walk is reset to an initial state in the case that it reaches threshold $m_i = N_i$.

Given that random walks may be activated in parallel, the transition probability rate must be modified accordingly. Previously, in the timer-only case, only a single transition was possible i.e. $S_{nm} = m_1 + 1$ and the total transition probability rate was determined by the rise-rate of this single random walk. Generalizing equation 2, the transition probability rate for an elementary random walk at time t is given by

$$w_i(t) = \frac{N_i}{T_i}, \tag{5}$$

where w_i is associated with random walk m_i. When multiple random walks are active in parallel, we compute the total transition probability rate as the sum of the individual transition probability rates associated with active elementary time-dependent random walks. Therefore, the total transition probability rate $W_n(t)$ is given by

$$W_n = \sum_{i \in A} w_i(t), \tag{6}$$

where A gives the index of active random walks.

In the model, changes in the rate at which the labile and nonlabile elementary random walks accumulate to threshold is subject to time-dependent changes. These rate modulations reflect the impact that visual changes sampled during a fixation have on the programming of an upcoming saccadic eye-movement. Two distinct types of random walk modulation are possible in the model, a rapid, fast-acting, surprise effect and a slower-acting encoding influence. The surprise effect is purely inhibitory

$$w_i(t) = \begin{cases} \alpha_i w_i(t) & \text{if } t \leqslant \beta_S \text{ and } 0 < \alpha_i \leqslant 1 \text{ and } i \in \{lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}, \tag{7}$$

for the UP condition and,

$$w_i(t) = \begin{cases} \epsilon_i w_i(t) & \text{if } t \leqslant \beta_S \text{ and } 0 < \epsilon_i \leqslant 1 \text{ and } i \in \{lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}, \quad (8)$$

where β and γ provide the temporal boundary during which modulation influences the random walk. In the case of a decrease in stimulus quality, encoding related inhibition of the random walk may occur. The onset occurs after the onset of surprise modulation and is given by

$$w_i(t) = \begin{cases} \zeta_i w_i(t) & \text{if } t \geqslant \gamma_E \text{ and } 0 < \zeta_i \leqslant 1 \text{ and } i \in \{lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}. \quad (9)$$

The transition probability rate is also subject to an increase under such circumstances where an increase in stimulus quality happens to occur. The scaling of the rate occurs in an analogous manner as for Equation 9 and operates on the same time-scale. However, the scaling factor λ now takes on values $\geqslant 1$ which effectively increases the transition probability rate,

$$w_i(t) = \begin{cases} \lambda_i w_i(t) & \text{if } t \geqslant \gamma_E \text{ and } 1 \leqslant \lambda_i < 2.5 \text{ and } i \in \{lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}. \quad (10)$$

4.4 MODEL FITTING

Optimization of model parameters was conducted by utilizing a genetic algorithm (GA) implemented in the Global Optimization Toolbox of MATLAB 2013b. The GA algorithm is an iterative procedure which simulates the behaviour of a set of candidate models (called a population) at each iteration (called a generation) of the algorithm. The behaviour of each candidate model is then assessed via an ob-

jective measure of performance called an objective function. At each generation, the parameters of a subset of models with the best performance according to the objective function are retained and used to seed the subsequent population.

Prior knowledge regarding plausible ranges for parameter values was supplied to the GA in the format of a set of constraints on the parameter values that could be used to generate a population of candidate models. Furthermore, given that the search space contains both integer (e.g. number of steps) and continuous valued (e.g. timer inhibition) a mixed-integer constrained search was used (Deep, Singh, Kansal & Mohan, 2009).

Simulation of the GA was conducted on the CCNi Distributed Matlab Cluster hosted at the University of Glasgow. The grid consists of 12 2x8 core Opteron CPUs each containing 128GB RAM running CentOS (rock clusters). We used the grid to concurrently estimate the fitness of 60 individuals per generations.

Each iteration of the genetic algorithm simulated fixation durations from 30 statistical subjects who viewed 100 arbitrary scenes. In each scene the statistical subject was presented with each condition a total of two times. Therefore, each trial yielded 6 critical fixation. From the individual data, fixation duration distributions were produced separately for each condition. Distributions of model predictions were produced by grouping the simulated data into 20 equally spaced bins.

The objective function that was used to assess candidate model fitness was defined as the likelihood $\ell(\theta\,;x_1,\ldots,x_n)$ of observing the empirical data $x_1 \ldots x_n$ given the parameters of the candidate model θ. Therefore, the optimization of this function by means of the GA constitutes a method of finding parameters that correspond to a maximum likelihood estimate $\arg\max \ell(\theta\,;x_1,\ldots,x_n)$ of the model given the data (DeGroot & Schervish, 2012).

For a single set of candidate parameters, a likelihood was calculated for each of the three conditions. Therefore, the value of the objective function was a sum of three likelihoods, one for each condition.

In the case that observations are independently and identically distributed, the likelihood function takes the following form

$$\mathcal{L}(\theta\,;\, x_1,\ldots,x_n) = \prod_{i=1}^{k} P(x_i|\theta). \tag{11}$$

Since the logarithm is a monotone function of the likelihood, maximizing the likelihood function is equivalent to maximization of the log-likelihood function

$$\operatorname{argmax}(\log(\mathcal{L}(\theta\,;\, x_1,\ldots,x_n))) = \operatorname{argmax}(\sum_{i=1}^{k} \log(P(x_i|\theta))). \tag{12}$$

Computing P for a specific observed fixation duration is equivalent to computing

$$\begin{aligned} \log(\mathcal{L}(\theta\,;\, x_1,\ldots,x_n)) &= \sum_{i=1}^{k} \log(q_i^{n_i}) \\ &= \sum_{i=1}^{k} n_i \log(q_i) \end{aligned} \tag{13}$$

where q_i is the probability of the i^{th} fixation determined by the model θ, and n_i is the number of times the i^{th} fixation duration was observed. In equation 14, n_i can be substituted for p_i which gives the proportion of the i^{th} fixation duration observed in the empirical data. This substitution is possible because $p_i > p_j$ whenever $n_i > n_j$ for all $i \neq j$. Since the GA optimizes a functional minimum value it is more convenient to instead to search for a minimum of the negative of the log-likelihood.

$$\operatorname{argmin}(-\log(\mathcal{L}(\theta\,;\, x_1,\ldots,x_n))) \quad . \tag{14}$$

4.5 RESULTS

4.5.1 *Empirical Data*

We used a dataset derived from Walshe & Nuthmann (2014) to empirically validate the adequacy of the model to account for eye-movements during natural tasks. Full details of this experiment can be retrieved from the original report (Walshe & Nuthmann, 2013, see also Chapter 2). To manipulate processing difficulty of the currently fixated stimulus, the present study employed a luminance manipulation. The change in luminance was used as a proxy for processing difficulty such that increases in luminance decreased the processing demands whereas decreases in luminance increased processing demands.

In Experiment 2 of this dataset participants viewed a total of 100 pictures of real-world scenes, in addition to 4 practice scenes. Each scene had a resolution of 800x600 pixels and was presented in full colour. Scenes were collected from online databases such as google images. They were selected to include a variety of categories such as indoor and outdoor as well as urban and nature scenes. Each scene was viewed by the participant only once over the duration of the experiment, and the experimental scenes were presented to the participants in a randomised order. Initially, scenes were presented at a baseline luminance of 60%. In order to observe the effect of relative luminance shifts on fixation durations, a luminance transformation was applied. Luminance shifted stimuli were created by converting the original scene into a $L*a*b$ colour space (Oliva & Schyns, 2000), and modifying the luminance channel L by the appropriate value. This procedure allows the separation of a luminance channel from the two colour channels, and permits the transformation of scene luminance independently of scene colour. Baseline and low luminance conditions constructed by a $L*.6$ and $L*.2$ transformation, respectively. The stimulus used in the high (100%) luminance condition was the untransformed scene.

Participants were instructed that they would take part in an experiment in which they would see many pictures of naturalistic content

and that their task was to encode the scenes for later recall. They were instructed that the recall phase would only begin once all the scenes had been viewed, but were not told how many scenes would be presented. These instructions were provided only to motivate scene encoding behaviour, and therefore the recall phase was not applied. Following the instructions, a nine-point eye-tracker calibration procedure was initiated. A trial began when the participant fixated on a cross presented at the centre of the screen. Following this fixation, the red cross and grey background were replaced with the scene presented at baseline luminance. Participants then engaged in the encoding task until a critical fixation was identified when a participant had made at least 10 saccades since the beginning of the trial. If a critical fixation had been identified, the luminance shift was made during the saccade immediately preceding the critical fixation. The luminance-shifted scene was presented for the entire duration of the critical fixation, and the luminance was then shifted back to baseline during the saccade immediately following the critical fixation. In total, four luminance manipulations were made on each trial; two manipulations resulted in an upward luminance shift, and two manipulations were made in the downward direction. After the first luminance manipulation had been completed, subsequent shifts occurred on every 10^{th} saccade since the most recent luminance shift. The order of the luminance shift direction (i.e. increase vs. decrease) was randomised within a trial. Once the fourth luminance shift had been made, and the participant terminated the resulting critical fixation, one second elapsed until the trial was terminated. The scene was then replaced with a grey background and red fixation cross. Once the participant fixated on the cross, the next trial was initiated. In the situation that the trial lasted longer than 25 seconds, the current trial was abandoned, and the participant was presented with a fixation cross to initiate the next trial. A schematic of the procedure for upward luminance shifts is presented in Figure 16.

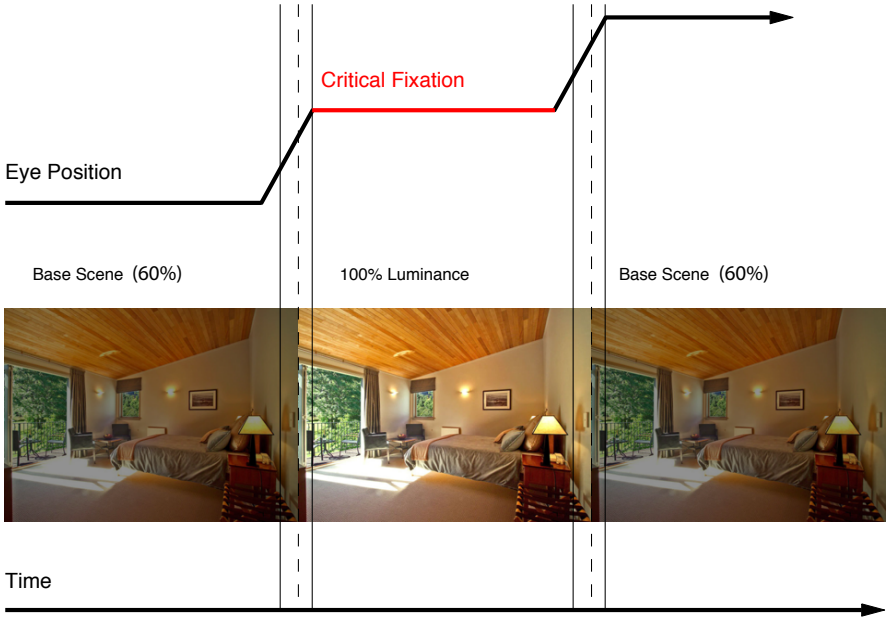

Figure 16: A schematic of the paradigm used to create gaze contingent luminance shifts. Base scenes (60% luminance) represent the image that is viewed during the fixation immediately preceding a critical fixation. A critical fixation is defined to occur on the 10^{th} fixation since the previous luminance manipulation. The oblique lines represent saccadic eye movements. During a saccadic eye movement, the scene is either increased or decreased in luminance. A critical fixation is terminated upon detection of a saccadic eye-movement, and the scene is restored to base scene luminance during this saccade.

Apparatus

Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz. The monitor screen was at a distance of 90 cm from the participant. During stimulus presentation, participants' eye movements were recorded using an SR Research EyeLink 1000 Desktop mount system. It was equipped with the 2000 Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz for each eye. Viewing was binocular, and both eyes were tracked. A chin rest was used in order to achieve stability of a participant's head position relative to the screen. The experiment was implemented in MATLAB 2009b using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner et al., 2007), which incorporates the EyeLink Toolbox extensions (Cornelissen et al., 2002). The software allowed precise control over the timing of display changes.

Online detection of saccades involves a speed-accuracy trade off, such that incorporating more samples reduces the noise in the signal. However, by increasing the number of samples, measurement lag is increased, which decreases the temporal precision with which saccades are detected. We implemented a 9-sample online velocity detection algorithm in MATLAB that aimed to mimic Data Viewer's offline saccade detection procedure (SR Research Ltd., 2006). Saccades were identified when gaze data from the right eye reached a two-dimensional velocity threshold of 85. Raw data was post-processed utilising SR Research Data Viewer to parse the gaze samples into sequences of fixations and saccades.

Analysis

Data were analysed with linear mixed-effects (LME) models, using the `lmer` programme of the `lme4` package (Bates et al., 2012) implemented in the R statistical computing software (R Development Core Team, 2012). To evaluate the effect of the downward and upward luminance shifts on fixation duration, we used treatment contrasts in which the baseline condition, where no luminance change occurred, served as the reference group. Consequently, the intercept for the fixed effect "luminance shift", estimates the mean value for the no-shift condition. The two slopes estimate the difference between downward luminance shift and no shift (DOWN) and between upward luminance shift and no shift (UP). The effect of luminance is assessed in the LME model by observing regression coefficients for the luminance shift conditions that are significantly different from 0; a two-tailed criterion of $t = 1.96$ was used to assess statistical significance. The LME models included random intercepts and random slopes for participants and items (Baayen et al., 2008).

PARTICIPANTS Four males and 18 females were recruited from the University of Edinburgh student population. The mean age of the participants was 21 years. Each participant was paid £7 per hour of participation in compensation for their time.

Results

The observed pattern of mean durations is plotted in Figure 17. In the LME model, the intercept for the fixed effect of luminance shift estimates the mean value for the no-shift condition ($b = 319.47, SE = 11.09, t = 28.79$). Downward luminance shifts were associated with critical fixations that were significantly longer than in the no-shift condition ($b = 124.28, SE = 13.15, t = 9.44$). In addition, a significant increase in fixation durations for upward luminance shifts ($b = 24.55, SE = 7.92, t = 3.10$) was observed. Relative to the no-shift baseline condition, fixation durations increased by 38.9% in the DOWN condition but only 7.7% in the UP condition.

Figure 18 shows the empirical distributions. In the DOWN condition, there is an overall shift towards longer fixation durations in both the UP and DOWN conditions. With regard to the tail of the distribution, in the DOWN condition there was a substantial increase in long fixation durations, which is manifested as a more positively skewed distribution. These distribution effects were formalized in greater detail in Walshe & Nuthmann (2014) and the reader can refer to this paper for a statistical validation of these qualitative effects on the distribution.

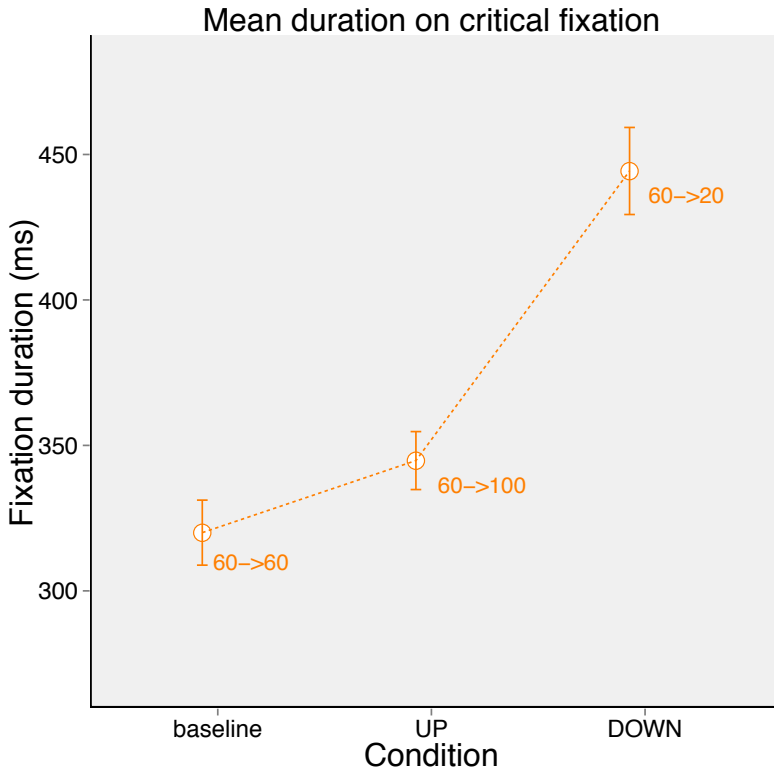

Figure 17: Mean fixation durations on critical fixations following gaze-contingent luminance shifts. Fixation durations are plotted as a function of the direction of luminance shift. Data is plotted for Experiment 1 (solid line) and for Experiment 2 (dashed line). Error bars represent the standard error of the mean.

4.5.2 *Model Analysis*

Model simulations were conducted to assess the sufficiency of the model architecture to account for observed differences in the empirically observed distributions. The parameters from the best fitting model are summarized in Table 9. The parameters that define the behaviour of the autonomous random timer are N_{timer} which is the number of steps in the random walk, and T_{timer} which specifies mean average timer duration. The number of steps in oculomotor random walks (i.e. labile, nonlabile, brain-to-eye lag, saccade execution) are given with $N_{oculomotor}$. The number of steps N and the mean timer duration T are used to calculate the transition probably rate of the random walks (see Equations 2 and 5). The mean duration of these stages is specified with T_{lab}, T_{nlab}, T_{BE}, T_{sack}. The param-

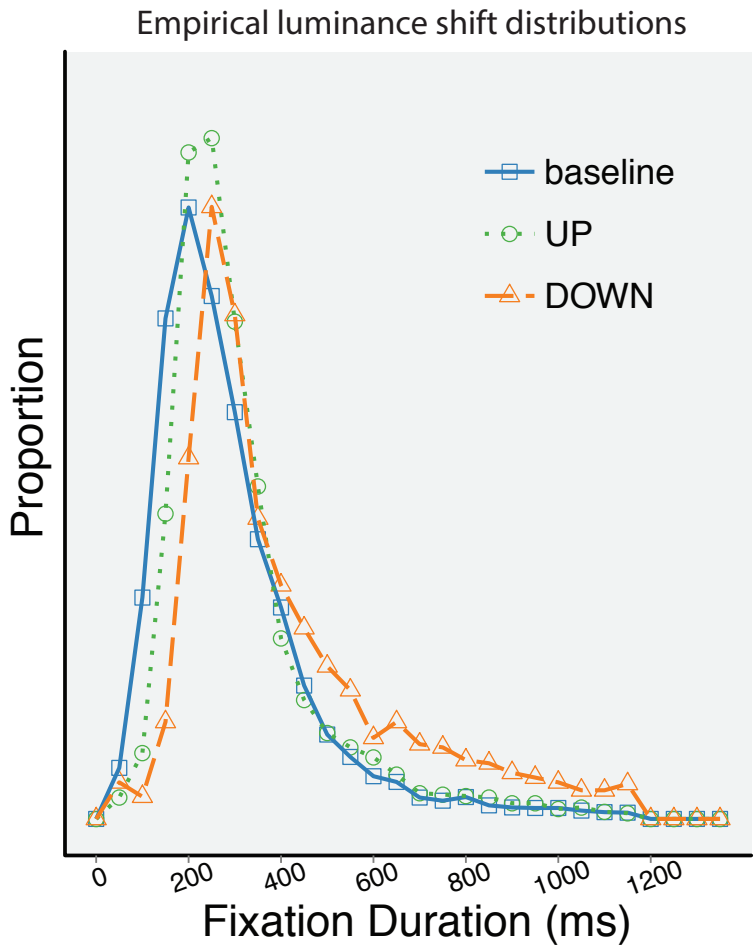

Figure 18: Fixation duration distributions. Empirical distributions for the three luminance conditions and their respective (ex-Gaussian) fitted distributions.

eters that determine the scaling of labile and nonlabile programming are α and ϵ and the parameters that govern the influence of encoding are specified with λ and ζ.

Baseline Simulations

We begin by assessing model behaviour in the baseline condition. All simulations use the same timer and programming parameters, but differ in whether a rate modulation is applied. For the baseline model, no rate modulation is applied. Baseline model parameters are found in Table 9. Proportional densities for the baseline condition are presented in Figure 19 and cumulative densities are presented in Figure 19. Proportional densities are simply the proportion of observations

Parameter	Symbol	Value	Equation	Range
Timer and Programming				
timer steps	N_{timer}	38.00	2	30 - 50
other steps	$N_{oculomotor}$	10.00	5	30 - 50
timer duration	T_{timer}	267.00	2	180 - 300
labile duration	T_{lab}	207.00	5	100 - 250
nonlabile duration	T_{nlab}	60.00	5	50 - 100
brain-to-eye duration	T_{BE}	30.00	5	Fixed
saccade execution	T_{sacc}	20.00	5	Fixed
Programming Modulation				
Up Condition				
up labile surprise	α_{lab}	0.81	7	.1 - 1
up nonlabile surprise	α_{nlab}	0.55	7	.1 - 1
facilitation labile	λ_{lab}	1.01	10	1 - 2.5
facilitation nonlabile	λ_{nlab}	1.19	10	1 - 2.5
Down Condition				
down labile surprise	ϵ_{lab}	0.76	8	.1 - 1
down nonlabile surprise	ϵ_{nlab}	0.93	8	.1 - 1
inhibition labile	ζ_{lab}	0.28	9	.1 - 1
inhibition nonlabile	ζ_{nlab}	0.25	9	.1 - 1
Modulation Offset and Onset				
surprise offset	β_S	98.27	7,8	25 - 125
encoding onset	γ_E	243.69	9,10	150 - 300

Table 9: UCM model parameters. The parameters were fit using a genetic algorithm with a maximum likelihood objective function. The N_{timer} is the number of steps in the random walk, $N_{oculomotor}$ is the number of steps in the oculomotor preparation random walks (e.g. labile, nonlabile). Mean random walk durations are specified with T_{timer}, T_{lab}, T_{nlab}, T_{BE}, and T_{sacc}. α,λ,ϵ and ζ specify the modification of saccade programming stages for the luminance UP and luminance DOWN conditions. β and γ specify the time course during which the rate modulation is activated.

found in each bin (bin size = 20). The figures reveal that the empirical data is well accounted for by the model with the best fitting parameters. The observed fixation durations for the baseline condition may be considered as a reference simulation to illustrate the main principles of the modelling. In effect, since the baseline condition does not include any specific influence on saccade programming due to scene changes, this models the case in which an individual is engaged in more or less naturalistic viewing behaviour (i.e. trying to remember

aspects of a scene). Simulation of the baseline data may be compared with a simulation study conducted in Nuthmann et al. (2010, see Simulation Study 1) which also models fixation durations during a similar naturalistic task.

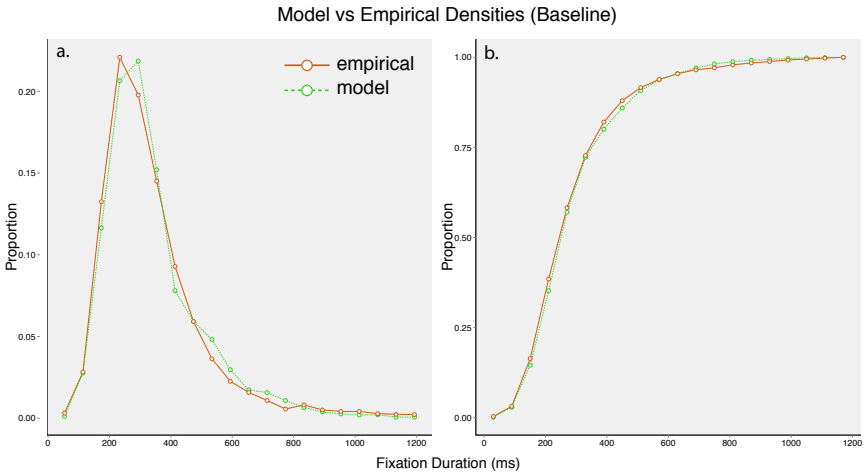

Figure 19: Model vs. Empirical Binned Densites. a) presents binned proportional densities, b) presents cumulative densities.

CANCELLATION Simulations using the baseline model reveal the role that saccade cancellation has in generating long fixation durations. A saccade cancellation occurs when the saccade timer attempts to initiate a labile saccade program while a labile program is already active. In this case, the old program is cancelled and is replaced by the new program. When cancellations occur within a fixation, the fixation duration tends to be longer than fixation with no cancellation. This consequence of saccade cancellation has been discussed in previous random timing accounts of fixation durations (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014, see Figure 4 and Figure 8 respectively). The extension in fixation duration is a consequence of the time required for a new saccade program to be initiated and executed to replace the previous program. It is also apparent that cancellations occur relatively infrequently compared to no cancellations. In the baseline condition, the proportion of no cancellation fixations is 78% , with one and two cancellations occurring on 17% and 4% of fixations. The remaining 1% of fixations contain three or more cancellations.

Figure 20 reveals that fixation durations in which no cancellation occurred result in a symmetric distribution. This is in contrast to the aggregate distribution which shows a strong positive skew (see Figure 19). The distribution of fixation durations in which 1 cancellation and 2 cancellations occurred are presented in the same figure. Grouping the distributions by cancellation reveals that fixation durations that include cancellations tend to be longer in duration than their no-cancellation counterparts. This is evidenced by a general shift in the mode of the distributions that occur as greater numbers of cancellations occur within the fixation[5]. Furthermore, these cancellation distributions are also symmetrical about the mean. It becomes clear from such a presentation that the overall skew in the distribution arises from a mixture of symmetrical distributions [6] each containing different proportions of the overall data. The component distributions arise naturally from the role that cancellation plays in extending fixation durations.

LABILE AND NONLABILE LATENCY Saccade programming occurs when a labile or nonlabile random walk processes are active. Therefore, the total time taken to program a saccade is equal to the sum of the durations of the random walks that eventually cause a saccade to be executed. Furthermore, the time required for the random walk to reach labile and nonlabile thresholds is anticipated to influence resulting saccade programs. The model architecture predicts that when saccade programming durations are relatively rapid, fixation durations should decrease and the converse should occur when saccade programming is slow. Figure 21 reveals the relationship between saccade programming and fixation durations. The figure presents fixation durations as a function of the sum of the labile and nonlabile latencies that generated the signal to move the eyes. Addi-

5 We present only one and two cancellation fixations for clarity. Increasing the number of cancellations included in the distribution results in greater rightward shifts in the distribution.

6 Symmetry in the distribution is not surprising. When there is no cancellation to interrupt the random walk growth process, the timing is equivalent to the sum of statistically independent draws from a exponential distribution. Under such conditions, the central limit theorem describes how values produced by such a process will converge to a normal distribution.

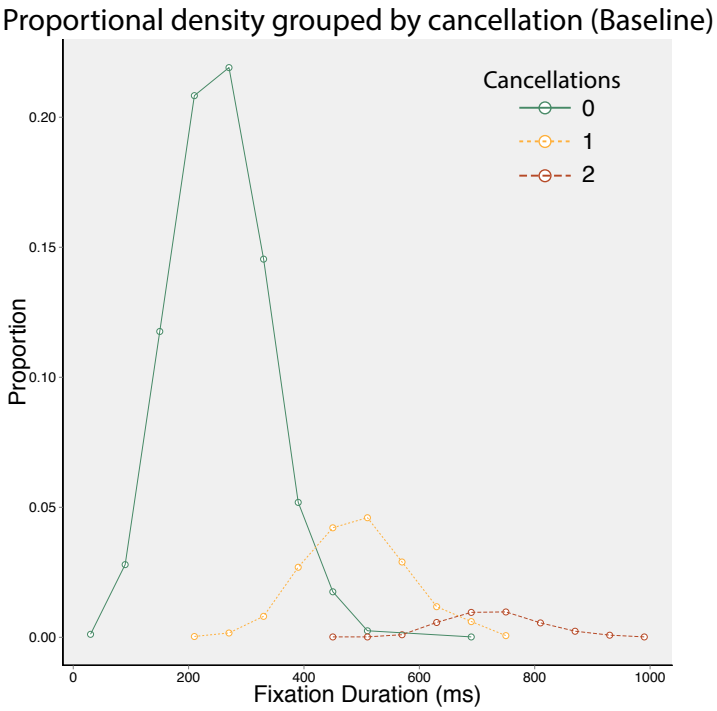

Figure 20: Model densities grouped by cancellation. The proportion of fixation durations observed in each bin grouped by the number of cancellations occurring in that fixation duration.

tionally, each data point is grouped by the number of cancellations that occurred during the fixation associated with the data point. It is evident from this plot that saccade cancellation results in a general increase in fixation durations. However, the figure also demonstrates that within each cluster the relationship between saccade programming and fixation duration follows a positive linear relationship.

ADVANCED PREPARATION OF SACCADES The time required to program a saccade is equal to the sum of the lengths of the labile and nonlabile programs. In the simplest case, a fixation onset occurs and then after some delay the saccade timer initiates a labile program. If there are no cancellations, then the fixation will terminate once the saccade execution stage of programming is initiated. However, in many cases an alternative scenario occurs. In these alternative cases, saccade programming has already been initiated by the saccade timer prior to the onset of a fixation. To measure how often this occurs within the baseline model we identified which stages of saccade pro-

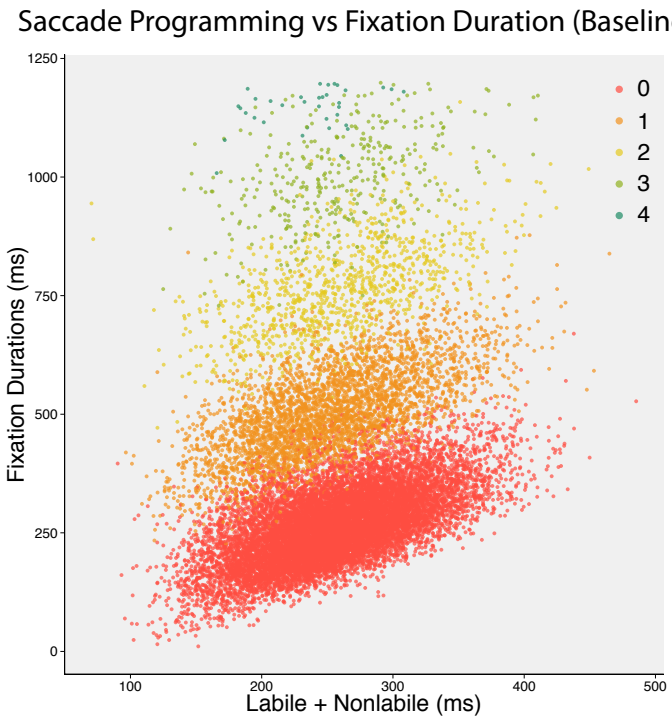

Figure 21: Fixation duration vs saccade programming. Each point represents a single fixation, clusters are determined by how many cancellations occured within a fixation. Fixation durations increase with longer latency saccade programming times.

gramming were active at the onset of a critical fixation. We found that at the time of critical fixation onset, in 49% of the cases a labile program was already underway and in $< 1\%$ of the cases a nonlabile program was engaged.

A consequence of pre-preparation of a saccade program prior to the onset of a fixation is a reduction in fixation duration. This occurs because of the link between saccade programming and fixation duration. Since a saccade has been partially prepared prior to the onset of fixation this implies a) no time is required for the saccade timer to initiate a labile program and b) labile (and possibly nonlabile) programming time is reduced. The consequences of pre-preparation are graphically demonstrated in Figure 22. This plot shows the amount of saccade programming that occurred prior to the onset of a critical fixation. As is expected, the more saccade programming is completed prior to the onset of a fixation, the shorter the fixation duration. A linear model predicts that for every 1 ms that programming was con-

ducted in advance, fixation times decrease by approximately 1 ms ($\beta = -0.95$ ms).

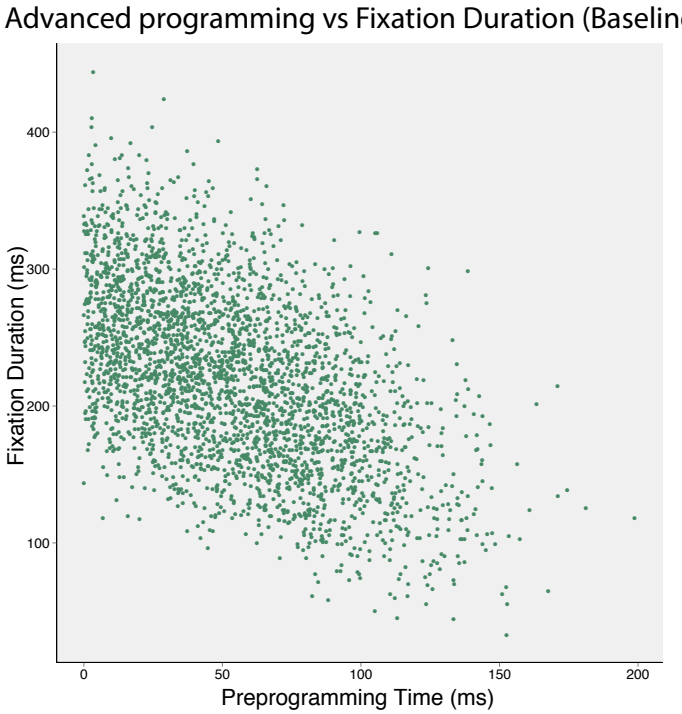

Figure 22: Advanced preparation predicts fixation durations. Fixation time decreases linearly with prior saccade preparation.

The analysis of the model in the baseline condition provides a window into it's basic mechanisms. Furthermore, it provides a grounding for understanding how specific influences that occur at the level of saccade preparation can come to influence the resulting distributions of fixation durations in characteristic ways. In the section that follows, we introduce how surprise and encoding modulation at the level of saccade programming interface with these basic mechanisms to explain eye-movement patterns in high-level scene viewing tasks.

Luminance Shift Simulations

Following up on the analysis on the analysis of the baseline simulations we now switch to simulations of the two luminance change conditions. The aim is to demonstrate how the timer rate modulations that occur in accordance with the surprise and encoding modulations as set out in the dual-process account combine to produce the observed distributional effects. Furthermore, we introduce a coun-

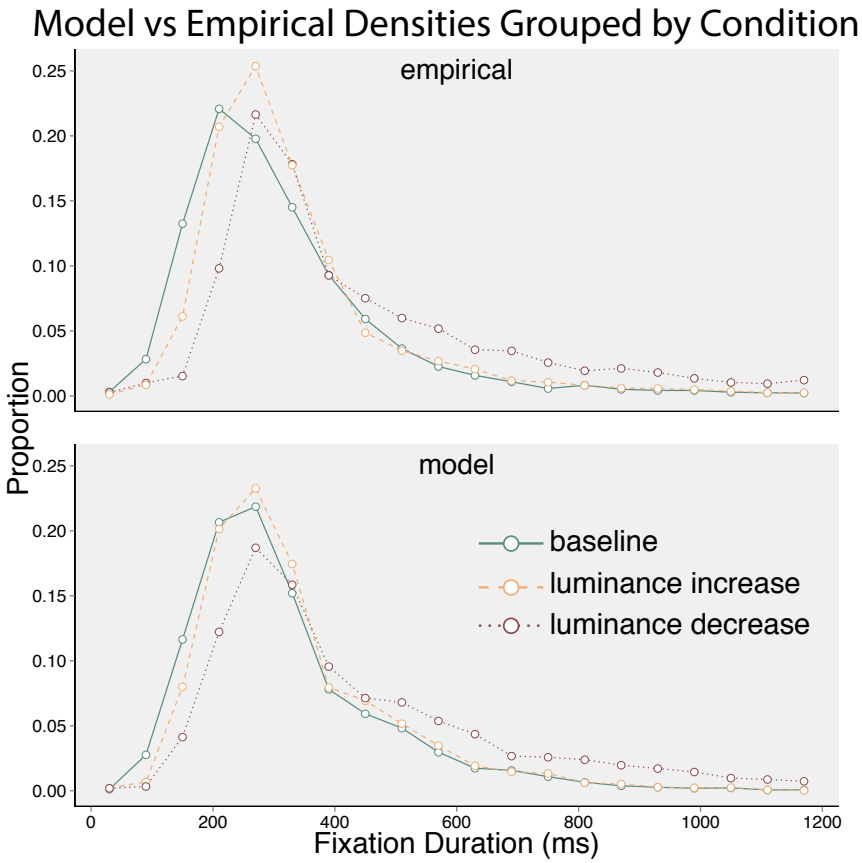

Figure 23: Model vs Empirical luminance distributions. The proportional densities are compared between humans and model predictions across the three luminance conditions.

terfactual analysis strategy to isolate the influence that each model component has on the aggregate distribution of fixation durations.

Cancellation

As in the baseline model, cancellation plays a large role in producing long tailed fixation duration distributions. Figure 24 shows the important role played by cancellation for each of the three conditions. Without cancellation, the distributions are largely symmetric and no long tail is apparent in any condition.

Analysis of the distribution in which cancellation has been removed illustrates that the rapid early shift in the distribution does not rely on cancellation. That is, after all cancellations have been removed from the distributions, a shift in the mean of the distributions still occurs. The mean of the baseline distribution is 254 ms while it is

264 ms and 285 ms in the DOWN condition. The fact that the shift in the distribution survives even after taking into account cancellation demonstrates that cancellation is not the model mechanism that produces this early shift.

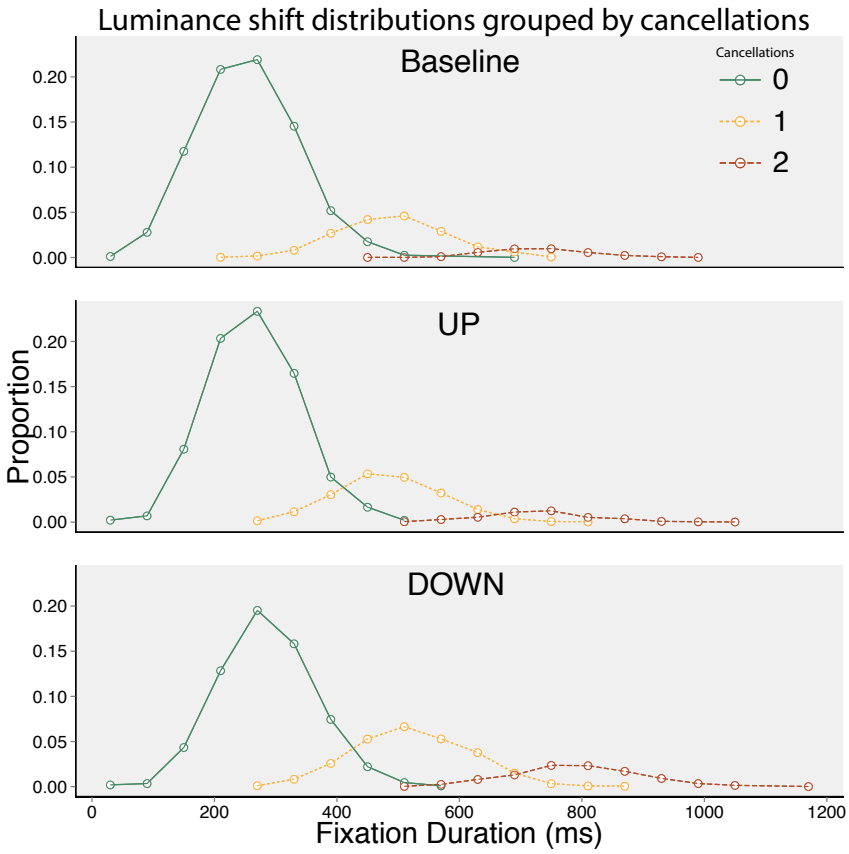

Figure 24: Model luminance densities grouped by cancellation. The proportional distribution of fixation durations varies as a function of the number of cancellations that occurred within the fixation. Cancellation distributions are shifted to the right and contain fewer observations. This is because cancellation extends fixation durations and are relatively rare compared to no cancellation fixations.

Figures 20 and 24 show the role that cancellation plays in producing long fixation durations. Therefore, it is plausible to consider that an increase in the density in the tail of the fixation duration distribution may partially arise from an increase in the rate at which saccade plans are cancelled. Analysis of the model behaviour does imply such a connection between cancellation and increases in long fixation durations. By analyzing the rate at which saccades are cancelled in the model it was determined that 22% are cancelled in the baseline con-

dition, 25% in the UP condition and this rises to 40% in the DOWN condition.

Counterfactual analyses

The preceding cancellation analysis reveals that an increase in the proportion of cancellations can account for specific increases in the tail of the distribution but cannot account for increases to very short fixation durations. However, to uncover the influence that the model mechanisms (i.e. surprise vs encoding modulation) have on observed fixation duration distributions, we adopt a counterfactual analysis strategy.

The logic of the counterfactual analysis is that we may compare the behaviour of full model θ with the behaviour of a model θ' that has had a single property ω removed. The comparison of these two models comprises a form of counterfactual comparison as we assume that behaviour that θ' produces is simply the difference between $\theta - \omega$ where we subtract out the model component associated with ω. We use this method of inference to isolate the causal impact that model mechanisms play in shaping the distribution of fixation durations.

COUNTERFACTUAL SURPRISE We begin the analysis by removal of the surprise process while leaving the encoding modulation process intact. Figure 25 reveals the impact of removing surprise in the UP condition (upper panels) and in the DOWN condition (lower panels). In these figures, the dashed line represents reference model θ while the solid line represents the counterfactual model θ'. The double-dashed line plots the baseline simulated data for comparison purposes. The most striking difference is the complete absence of any impact on very short fixation durations. Removing the surprise in the UP condition results in a distribution that almost perfectly matches the baseline distribution. This implies that the early shift in the mode is caused by the early-onset surprise process that inhibits timer growth. This inference is warranted by the fact that no influence on short fixation durations survives after the removal of the surprise influence.

Removing surprise from the DOWN condition has a more nuanced influenced on the distribution. Like the UP condition, surprise has a causal role in temporally extending short fixation durations. Evidence for this arises from the fact that there is very little difference between the baseline and surprise-absent model for short fixation durations. However, removing surprise does not eliminate the effects that occur in the tail of the distribution. The logic of the counterfactual analysis implies that the increase in long fixation durations are not causally explained by the surprise effect. If they were, it would then be expected that the counterfactual model and the baseline model should be at equality for long fixation durations as well.

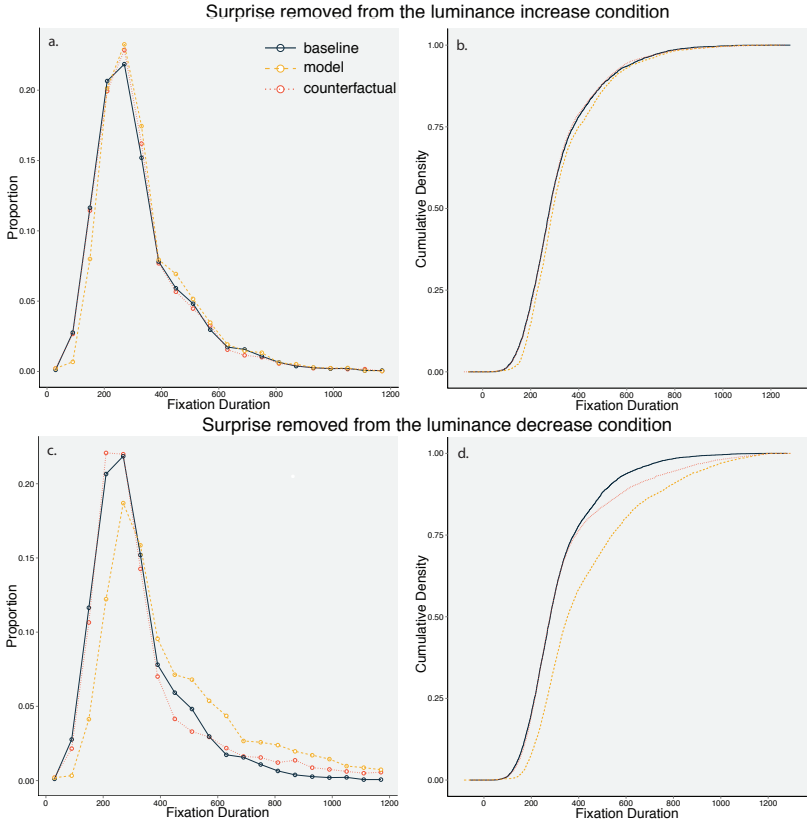

Figure 25: Counterfactual Surprise Analysis. Each plot contains a comparison of the distribution from the model baseline condition, the model luminance shift condition and the model counterfactual in which surprise is removed from the model simulation. Proportional densities from the luminance increase condition are plotted in A and CDFs are plotted in B. In C the proportional densities are plotted from the luminance decrease condition and D plots CDFs from this condition.

COUNTERFACTUAL ENCODING MODULATION The preceding analysis used a counterfactual approach to illustrate that the early onset surprise effect is causally implicated in the shift that occurs for short fixation durations. Furthermore, it was demonstrated that this surprise effect cannot account for an increased prevalence of long fixation durations. The analysis now utilizes the counterfactual strategy to illustrate the selective role that late encoding modulation influences have on long fixation durations. Figure 26 shows how the model behaves when encoding influences have been removed. Particularly informative is the behaviour of the counterfactual model in the luminance decrease condition (lower panels of Figure 26). Here, encoding modulation has been removed from the model but the surprise effect remains intact. This counterfactual analysis represents an inverse to the case in which surprise is removed but encoding is retained. As such, an inverse pattern is observed. For short fixation durations, removing encoding modulation results in very little difference between the counterfactual and full model. This fact illustrates that encoding modulation is not necessary to produce early shifts in the distribution. That is, removing encoding modulation has no impact on the early shift in fixation durations that are observed. A corollary of this is that surprise alone is fully capable of producing the early shifted distribution. However, for long fixation durations removal of the inhibitory influence of encoding modulation results in a reduction in the proportion of long fixation durations. This can be seen by inspecting Figure 26 (lower panels) where a reduction in long fixation durations is observed. Counterfactual inference implies that encoding inhibition specifically influences long fixation durations.

Counterfactual analysis of the UP condition reveals that encoding modulation plays very little role in shaping the distribution in this condition (upper plots of Figure 26). This behaviour is to be expected as the best fitting model parameters were very close to unity (i.e. no influence). It should also be noted that counterfactual removal of encoding modulation does not influence the early shift in the distribution. This is consistent with the observation from the DOWN condition that surprise and encoding modulation can be dissociated.

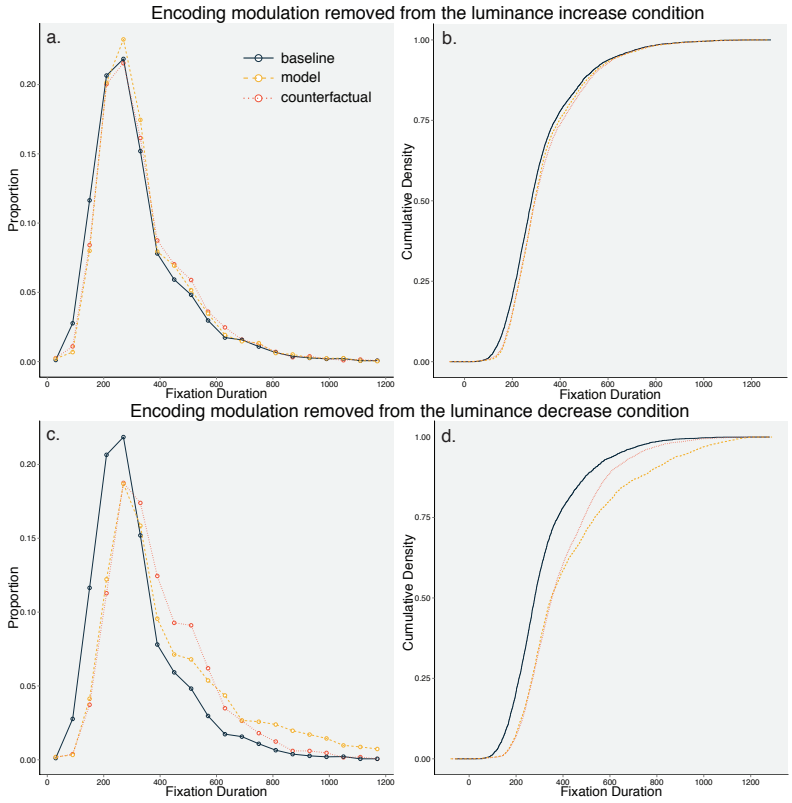

Figure 26: Counterfactual Encoding Analysis. Each plot contains a comparison of the distribution from the model baseline condition, the model luminance shift condition and the model counterfactual in which encoding is removed from the model simulation. Proportional densities from the luminance increase condition are plotted in A and CDFs are plotted in B. In C the proportional densities are plotted from the luminance decrease condition and D plots CDFs from this condition.

Analysis Summary

In summary, the model analysis reveals the mechanisms that combine to produce a) distributions that occur under ordinary baseline conditions b) changes to those distributions arising from changes to the quality of stimulus content. Changing the quality of the stimulus results in large scale changes in the distribution of fixation durations observed in the experiment. These changes are consistent with observations from a number of authors who have suggested that changing the stimulus quality in such a fashion results in a rapid influence on fixation durations (surprise) and a slower acting influence (encoding modulation) which we have labelled the dual-process account (Glaholt et al., 2013; Walshe & Nuthmann, 2014; Henderson

et al., 2014). Specifically, the UCM model shows how an inhibitory surprise mechanism that slows down the rate at which saccades are programmed directly influences even the shortest fixation durations. Counterfactual simulations were used to show how removal of this surprise process completely negates the influence on short fixation durations. Counterfactual simulations were also used to illustrate the role that later encoding modulations have on distributions of fixation durations. We showed that counterfactual removal of encoding modulation resulted in no influence on short fixation durations but had a considerable influence on long fixation durations only.

4.6 GENERAL DISCUSSION

Decision making may be thought of as a process that evolves over time, and in certain cases, results in behavioural outcomes that can be measured in precise ways. Eye-movements are one type of behaviour that may be considered from this perspective. The primary goal of an eye-movement is to select a behaviourally relevant source of stimulation. During any single fixation, the visual system must choose which, out of many possible options, to make the target of the subsequent eye-movement (Schall, 2001). A simple question that may be asked is, what factors influence the timing of such a decision?

4.6.1 *Evidence for surprise and encoding influences during naturalistic scene viewing*

The model that has been introduced here combines principles that have been developed to explain patterns of eye-movements in reading (Reichle et al., 1998; Engbert et al., 2005; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), visual search (Trukenbrod & Engbert, 2014), and scene perception (Nuthmann et al., 2010). The core principles developed in the UCM model bear most resemblance to recent random-walk timer based accounts of eye-movement decision making (Schad & Engbert, 2012; Trukenbrod & Engbert, 2014; Nuthmann

et al., 2010). While many differences exist in the way these models are implemented, they form a class. The family resemblance derives from a) their implementation of eye-movement decisions as arising from a sequential sampling process, b) saccade programs are triggered by a rhythmic timer c) multi-stage saccade programming architectures. Rather than presenting a competing account of eye-movement control, the modelling results presented here complement recent efforts by showing that the basic principles of this mechanism can generalize to novel task domains. Furthermore, the simulations argue that additional assumptions that have not been made by other models are required to account for recently observed empirical results in the scene-viewing literature.

Recent empirical work studying eye-movements during naturalistic viewing conditions has pointed to the existence of two distinctive types of influences on the timing of eye-movement decisions. Glaholt et al. (2013) provided a detailed analysis of how eye-movements adapt to scene-quality changes that are made during a saccadic eye-movement. First, they observed that manipulating a scene by removing spatial-frequency content during a saccadic eye-movement results in an immediate increase in fixation durations. This result provided additional support for the hypothesis that fixation durations during scene viewing may be directly controlled by stimulus content (Henderson et al., 2013; Henderson & Pierce, 2008). Glaholt et al. (2013) provided additional analyses based on the shape of fixation duration distributions. They discovered a pattern in their results showing two distinctive distributional effects: an influence on the central tendency and an influence on the tail. They found that both low- and high-pass filtering resulted in a shift in the central tendency of the distributions. However, they also found that low-pass filtering resulted in a larger influence on the tail of the distributions. They argued that the removal of high-spatial frequencies induces greater challenges to scene encoding processes than does removal of low-spatial frequencies alone. Interestingly, this influence on the tail of the distribution was no longer present when the entire scene was flipped vertically

or horizontally suggesting that such a large scale modification of the scene eliminates any high-pass filtering benefit.

In a later study, Walshe & Nuthmann (2014) found that when the luminance is decreased during a saccade, fixation durations are increased. Analyzing the distributions in this study revealed a pattern consistent with that reported by Glaholt et al. (2013). When the reduction in luminance was great enough, an increase in both the central tendency as well as the tail was observed. Interestingly, fixation durations were also found to increase when the luminance of the scene was increased. When luminance was increased it was shown that the increase in fixation durations primarily came from an increase in the central tendency and there was no specific change in the tail of the distribution.

Both Glaholt et al. (2013) and Walshe & Nuthmann (2014) suggested an explanation for these results that may be referred to as a dual-process account. Specifically, it has been suggested that such direct influences on fixation durations in scene viewing may arise from two possible sources. The first is a surprise influence. It is hypothesized that this mechanism arises due to sudden, unexpected visual changes. An important aspect of surprise is that it does not necessarily depend on stimulus complexity or higher order features. Therefore, it can be triggered by detection of simple visual changes across saccadic eye-movements. Thus, the surprise effect may occur very rapidly. In contrast, influence on the tail of the distribution was speculated to arise once a more detailed level of analysis has been conducted on the stimulus. This late, encoding related, influence on the distribution occurs systematically in instances in which the stimulus changes are likely to result in additional processing difficulties. On this basis it was hypothesized that, complementing the early surprise effect, a late onset encoding related influence on fixation durations is liable to occur.

The present modelling efforts represent a formalization of this hypothesis. The model extends previous models of eye-movement control in high-level tasks by assuming surprise-related and processing-related modifications to the duration of saccade programming sig-

nals. Therefore, the present model provides a formal hypothesis for the mechanisms underlying recently observed direct-control effects in scene-viewing and contributes to a growing body of evidence suggesting that eye-movement decisions in high-level tasks may be accounted for by the accumulation of decision related signals (Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014).

4.6.2 *Relationship to the CRISP and ICAT models*

The UCM model has substantial overlap with previous models of eye-movement control in high-level tasks, while making certain novel assumptions. In the CRISP model, a timer initiates saccade programs at random intervals that is not directly tied to fixated content. Thus, the saccade timer creates a partial decoupling of the initiation of saccade programming from stimulus processing. For example, the saccade timer may can initiate a saccade program immediately after fixation onset. In this case a saccade program is initiated that may terminate the fixation but very little, if any, stimulus processing has occurred. This situation is more extreme when the saccade timer initiates a saccade program prior to the onset of a fixation. In these cases, if the saccade program is not cancelled, the saccade program that will eventually terminate the fixation was initiated on the previous fixation. However, the decoupling of the saccade programming and saccade timing is only partial. The saccade timer may be inhibited in the case that processing difficulty is high. This inhibition can result in delays in the time required to initiate saccade programs and therefore can result in extensions in fixation durations. This, along with saccade cancellation, is the primary mechanism of direct control in the CRISP model. Importantly, the CRISP model does not permit any stimulus dependent changes in the duration of saccade programming (i.e. labile and nonlabile).[7]

Direct control in the UCM model operates in a distinct manner from the CRISP model. Like the CRISP model, the UCM model initi-

7 Although no systematic changes in the duration of saccade programming are permitted, labile programs can be cancelled.

ates saccade programs according to randomly timed intervals. However, the UCM implements the random timer as an autonomous timing signal that is fully decoupled from stimulus processing. In other words, no stimulus related influences can interfere with the rate at which the timer initiates saccade programs. Thus, the random timer in the UCM is an instance of a fully indirect control mechanism. In the UCM model, direct control operates at the level of saccade programming. Processing related changes modify the rate at which stages of saccade programs are completed. This is consistent with basic oculomotor research suggesting that saccade programming time can reflect the difficulty of the saccadic decision making task (Carpenter & Williams, 1995; Camalier et al., 2007; Smith & Ratcliff, 2004).

The difference in direct control assumptions between the UCM model and CRISP has important implications. A consequence of the direct control mechanism implemented in the CRISP model is that when a saccade has been partially prepared prior to the onset of a fixation the stimulus content present within that fixation may play no role in governing the duration of saccade programming. This is an important assumption for the CRISP model in that simulations have shown that up to 72% of fixations are expected to have a labile program already activated when the fixation begins (see Table 4 Nuthmann et al., 2010). Furthermore, when fixation durations are short, saccade programming also tends to have been completed prior to the onset of the fixation (see Figure 4 Nuthmann et al., 2010, also see present results, Figure 22). Therefore, in the CRISP model, the only method by which fixations with advanced prepared labile programs may be influenced by stimulus content is through the cancellation of a labile program. In the UCM model, labile programs initiated prior to the onset of a fixation may be extended in time. This model assumption allows even very short fixation durations to be directly controlled by stimulus content.

Unlike the CRISP model, ICAT produces stimulus dependent changes to the rate at which saccade programs are initiated (i.e. saccade timer) and in the duration required for a labile program to complete. Thus, increases in fixation durations can arise from delays in the initiation

of a saccade program or, more directly, in the time required for saccade programming to complete. Furthermore, since ICAT permits the duration of labile programs to be modified on a moment-to-moment stimulus dependent basis, it accounts for adjustments in the duration of saccade programs that were initiated on a previous fixation. For this reason, ICAT can, in principle, account for increases in very short fixation durations. However, ICAT does not assume any distinction between surprise or encoding related influences on the initiation or programming of saccades. As such, it is difficult to see in what way ICAT would predict differential effects on the mode and tail of the distribution that have been observed in scene viewing tasks (Glaholt et al., 2013; Walshe & Nuthmann, 2014).

4.6.3 *Inhibition of saccade programming*

A novel assumption in the present modelling is the introduction of surprise and encoding modulation signals that modify the rate of saccade programming at different time scales. The surprise signal modifies the transition probability immediately after the eye-brain lag and lasts for 98 ms. The encoding signal only begins to influence fixation durations after 243 ms and lasts until a saccadic eye-movement is made. Analysis of the simulations revealed that the surprise signal is crucial for producing the early shift in the distribution but plays very little part in extending very long fixation durations. We used a counterfactual method of analyzing the influence of a specific model component. The logic of this method involves the assumption that two models are identical in all aspects except for one critical feature, for example, surprise. We claim that by assessing the counterfactual model, any difference in the distribution between the two models must arise from that one single model component. In Figure 25 **c** and **d** it is clear that by removing surprise, the early shift in the distribution is nearly completely removed. However, removing surprise does not eliminate the effects present in the tail of the distribution. On the other hand, Figure 26 **d** gives an indication of the influence that encoding modulation plays in determining critical aspects of the

fitted distribution. Removing encoding modulation had very little influence on the early parts of the distribution. The early shift, that it is argued arises due to surprise, remains. However, the removal of encoding modulation results in a clear reduction in the proportion of long fixation durations. This illustrates the targeted influence that encoding modulation has on the tail of the distribution. When encoding modulation is removed, what remains is the general, and rapid onset, increase in fixation durations due to surprise.

4.6.4 *Facilitation of saccade programming*

At the level of encoding, the model contains parameters that may be used to increase the rate of accumulation. As may be expected, in the modelling of the luminance data, these parameters played very little role in determining the shape of the predicted fixation duration distributions. The best fitting facilitation parameter increased the rate of labile saccade programming by only 1% and nonlabile programming by 19% during nonlabile programming. This may be contrasted with the inhibitory influence which decreased the rate by 72% in the labile and 75% in the nonlabile stage of programming. Figure 26 **a** and **b** provide a graphical demonstration of the relative lack of influence of encoding facilitation in the present experiment. Both counterfactual and model simulations are very similar.

Based on these simulations, there is no indication that facilitation is required to model fixation durations arising from an increase in luminance. However, there is some limited evidence that such bidirectional modifications to fixation timing does exist. In a gaze-contingent scene-quality approach using spatial-frequency filtered scenes, Henderson et al. (2014) reported that fixation durations decreased following a scene change in which the scene went from low-pass to unfiltered. Therefore, these results imply that facilitation does play a part in determining fixation times in scene viewing. However, as has been discussed in Section 3.3 this result has yet to be replicated. Therefore, we have included a facilitation parameter in the present modelling as a generalized hypothesis about how fixation timings are set. More

experimental work is clearly required to determine which stimulus conditions elicit facilitation.

4.6.5 *Predictive model of surprise modulation*

Surprise related changes in the timing of saccadic eye-movements may be understood within the context of a predictive models of perception and action. Within this perspective, the surprise related changes in the planning of an impending eye-movement results from the mismatch between the statistics of the visual environment that are expected to be present prior to an eye-movement, and the statistics (i.e. lower luminance environment) that are discovered at the onset of the next fixation. To the extent to which these expectations are violated, the system responds by reevaluating the motor control parameters that are used to govern the timing of saccades. In studies of saccadic reaction times it has been shown that latency increases with decreasing target probability (Carpenter & Williams, 1995). This fact has been used to suggest that saccade related decisions are impacted by the probabilistic context in which a target is embedded. It has also been shown that saccadic reactions increase with the degree to which the location of a target violates prior expectations (Vossel, Mathys, Daunizeau, Bauer, Driver, Friston & Stephan, 2014). The large-scale stimulus change that occurs in such gaze-contingent scene quality studies may induce such a violation of expectations. In predictive coding accounts of perceptual processing, it is claimed that perceptual circuits convey signals relative to the difference between anticipated and observed inputs (Huang & Rao, 2011). One functional significance of such a coding system is that redundancies in perceptual inputs may be reduced or removed, allowing for a more compact and efficient neural code. Predictive signals may also be used as a method of smoothing errors and deemphasizing perceptual samples arising due to chance fluctuations (Burr & Cicchini, 2014). However, an adaptive system must also retain the capacity to adjust its behaviour when the sensory evidence suggests a large-scale violation of expectations. Such an error signal is used by the motor system to adjust saccade

amplitudes in the context of saccadic (Wallman & Fuchs, 1998) and motor adaptation tasks (Berniker & Kording, 2008) where the motor gain on future trials depends on the observed error in the current response. Predictive signalling has been shown to exist as early as the retina (Hosoya, Baccus & Meister, 2005) and at later stages of the visual pathway including LGN (Dong & Atick, 1995) and visual cortex (Rao & Ballard, 1997). In the present context, it is possible that the large scale visual change in the structure of the image induces a change in the spatial structure which leads to an error signal being generated. This error signal may rapidly reach FEF or SC resulting in a modification to saccade planning, which is here modelled as a slow-down in saccade related decision signals.

4.6.6 *Future Work*

Future modelling work in this direction should aim to test the principles developed here within a wider range of experimental contexts. Distributional patterns in which an early shift and late influence on the tail are well accounted for by the dual-process model introduced here. A more complete analysis of this behaviour would be to show, in greater precision, how parametric changes in surprise and encoding related scene changes translate into model parameters. For example, in the current formulation an increase in the surprise parameter will translate into greater inhibition of saccade programming rates and therefore longer fixation durations. Walshe & Nuthmann (2014) provides evidence that surprise does scale in such a way. In their study, a larger change in luminance elicited larger changes in surprise related fixation durations. Future studies could be conducted to attempt to map out whether fixation durations scale parametrically with surprise, as is predicted by the UCM model.

The status of the role of facilitation in the model also requires further elaboration. Currently, it plays very little role in producing the fitted distributions. Henderson et al. (2014)'s data suggests that facilitation may occur, but may be masked in many cases by the surprise that is elicited by the large scale visual change. This interpretation is

strengthened to some degree by studies reported in this thesis. Section 3.3 and 3.4 both show this trend, in that greater stimulus clarification results in lower costs on fixation durations. This occurs despite the fact that greater stimulus clarification also results in a more visually distinct image. The inability to replicate Henderson et al. (2014)'s spatial frequency study precludes strong interpretation of this data. Nevertheless, tentative results indicate that additional studies exploring the masked-facilitation hypothesis are warranted. A prediction implicit in the current modelling is that masked facilitation can arise due to interactions between the surprise and encoding facilitation mechanisms.

4.6.7 *Conclusion*

This chapter provides a hypothesis for how the eye-movement control system adapts to stimulus changes that occur on a fixation to fixation basis. The novel hypothesis that we introduce is that there are two qualitatively distinct types of influence, a rapid onset surprise and a delayed onset encoding modulation. It is shown that this hypothesis is capable of explaining fixation duration distributions derived from luminance changes, and it is claimed that this pattern can be generalized to explain other recent results reported in the literature. Future work must be conducted to assess the generalizability of these simulations to novel task domains and stimulus variants.

5

THESIS SUMMARY

Eye-movements have been, and will continue to be, a behaviour that captures interest from all areas of cognitive science. Firstly, they are one of the most frequent behaviours that humans make. Each day, we shift our eyes from one location to another more than 100,000 times. The visual system must make each of these gaze changes with care in order to extract the most relevant visual information possible from the environment.

The work in this thesis has concerned itself with the question of the *timing* of these eye-movements. How does the brain decide whether it should spend more, or less, time looking at a specific location. As we have shown, this timing is not random - it adapts to the visual processing environment. The point of departure for our studies of eye-movement timing began with a study using a double-step procedure. The power of the double-step procedure derives in part from its simplicity. The movement of two simple targets can give rise to extraordinarily complex patterns of results. However, the simplicity of the experimental design makes it possible to infer what mechanisms are causing this behaviour without making overly elaborate assumptions.

Typically, studies of double-step experiments have been conducted in highly constrained environments, such as dots flashed on blank backgrounds. This is with good reason. When the goal is to study the fundamental nature of basic mechanisms it is important to design a task that measures the behaviour as efficiently as possible (i.e. little noise). However, one problem with this approach is that human vision did not adapt to process targets in these environments. Rather, the evolution of the human visual system has been designed to process information when it is embedded in natural scenes. In comparison to the static backgrounds used in basic tasks, natural

scenes are rich with structure and patterns. This difference leads to two questions. First, we may ask whether the behaviour that occurs in the basic task happens at all when taken into a natural setting. Second, we may also ask whether there are characteristic patterns of behaviour that occur in the natural setting that do not occur in the basic setting. In both cases, these are issues of generalizability. Our double-step task contributes to both of these concerns. Firstly, we have confirmed that saccades to double-step targets are modified in an analogous way in natural scenes and in static, black background, conditions. Importantly, we show that saccades in black backgrounds and in natural scenes differ in important ways. A measurement of saccade behaviour known as *saccadic dead time* is the minimal amount of time that a target must be presented if it is going to have any influence on modification of the upcoming saccade. We show that this important measure of saccade behaviour differs depending on whether the saccade is being programmed to a target in a natural scene as opposed to a non-natural scene. This study highlights the role that naturalistic stimuli have on saccade behaviour.

The relationship of Part I to Parts II and III arises due to the important role that double-step tasks have had on the foundational assumptions of models of eye-movement control in high-level tasks. The double-step task used in Chapter 1 is used to study how saccades are modified in response to new sources of information. However, the double-step task has also been used to determine that saccades can be programmed in parallel. Both of these assumptions have been utilized in the modelling that was conducted in Parts III on the data reported in Parts II.

Part III represents an effort to synthesize empirical results observed in Part II as well as from other authors Glaholt et al. (2013); Henderson et al. (2014) into a formal, testable model. The model owes it's main theoretical basis to an account that we have dubbed the *dual-process account*. This account was first formulated in (Glaholt et al., 2013), and the ideas presented within the modelling section owe a great deal to the those authors' careful examination of surprise and encoding modulation effects. The simulations show that by using

a framework that has considerable conceptual similarity to previous models of eye-movements in high-level tasks (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014), that we are able to explain the main pattern of results observed in Chapter 2. Although we only simulate one dataset, we claim that this pattern can generalize to other results in Chapter 2 as well as those reported by other authors (Glaholt & Reingold, 2012; Henderson et al., 2014).

In summary, the thesis aims to uncover the mechanisms of eye-movement control from multiple levels of abstraction. At the most basic level, the thesis utilizes a double-step procedure to gain a high-resolution view into the operation, and timing, of basic mechanisms. Abstracting from these basic processes, we used gaze-contingent methods to observe how fixations respond to varying the moment-to-moment quality of scene stimuli. Finally, by utilizing computational modelling, we gain an understanding of how these processes fit together into a mechanistic account of eye-movement control in high-level tasks.

Part IV

APPENDIX

A

APPENDIX A

Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model

R. Calen Walshe (r.c.walshe@sms.ed.ac.uk)
Antje Nuthmann (antje.nuthmann@ed.ac.uk)
University of Edinburgh, Psychology Department,
7 George Square, Edinburgh EH8 9JZ

Abstract

Several computational models explaining fixation durations in scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010) and in reading (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998) assume that saccade programming is completed in two stages: an initial, labile stage that is subject to cancellation and an subsequent, non-labile stage in which the program can no longer be cancelled. This distinction is motivated by findings from double-step experiments that used much simpler situations than scene viewing or reading. Here, we adopt a classic double-step paradigm to a scene-viewing context. In a Static condition targets are presented to the left or right of a central fixation cross along a horizontal axis while in a Scene condition targets are presented in a gaze contingent manner along a trajectory defined by the location of recent fixations. We found evidence in support of the claims that saccade cancellation occurs within a naturalistic scene-viewing context and that saccade cancellation can account for increases in observed fixation duration distributions. The duration of the non-labile stage was estimated to be longer in the Scene condition compared to the Static condition.

Keywords: Double-step; Scene viewing; Saccade programming; Mixed-effects modelling

Introduction

There is a long history of utilizing the double-step paradigm to explore the lower level details of the programming and execution of eye-movements (Westheimer, 1954). Classic variations of the double-step paradigm involve presenting participants with two targets along a horizontal axis with a varying inter-stimulus interval separating the two targets. For instance, in one classic study of saccade programming that utilized double-step stimuli, Becker and Jürgens (1979) had a condition in which a first target was presented at 15° to the left or right of fixation with a second target presented at 30° in the same direction at delays of 50, 100, 150 and 200 ms. The participants task was to fixate the target as quickly as possible, which meant that in order to fixate the more distal target, a saccade program initiated to the first target was put in competition with a program to the second target.

By studying the characteristics of the response pattern, the paradigm affords numerous avenues to investigate the processes underlying the programming of saccades. One method of formalizing double-step data has been to produce what is called an amplitude transition function (ATF) (Becker & Jürgens, 1979). The ATF provides a measure of the saccade amplitude resulting from the stimuli as a function of the delay (D) which measures the time elapsed between the onset of the second target step and the first measured response saccade. Therefore, in this analysis only those trials in which both targets appeared prior to the first response saccade are analysed. D can therefore be thought of as the amount of time available to the saccadic system to reprogram an eye-movement to the second target. Frequently replicated results demonstrate that when reprocessing time in low (short values of D) saccades are typically directed towards the first target step, and when reprocessing time is high (high values of D) then saccades compensate for the updated target position and move to the second target step (Ludwig, Mildinhall, & Gilchrist, 2007).

From inspection of the ATF it is also apparent that there is a point at which the appearance of the second target step can no longer have an influence on saccade programming. This region of the distribution represents a "point of no return" in the saccade processing to the first target and as such the program to the first target is executed despite the availability of countermanding information from the second target. The point in processing at which a saccade program can no longer be modified by a second target is also referred to as saccadic dead time (SDT) and has been estimated at approximately 80 ms prior to the execution of a saccade (Ludwig et al., 2007).

The double-step paradigm has been a fruitful one in elucidating the basic properties of the occulomotor system. The principles derived from such investigations have formed the basis of several models of eye-movement control in a variety of fields. These investigations have proven particularly useful in models that attempt to explain the mechanisms that control how long aspects of the visual environment are fixated. For instance, Nuthmann et al. (2010)'s CRISP model which explains fixation durations in scene viewing, utilized a two stage saccade programming mechanism. In the first labile stage of programming a saccade could be cancelled and reprogrammed, while a program that had moved into the non-labile stage could no longer be cancelled. In the CRISP model architecture saccade cancellation acts as a causal mechanism that accounts for systematic delay in fixation durations. The theoretical dichotomy between a labile and nonlabile stage of programming was first introduced in Reichle et al. (1998)'s E-Z Reader model of eye-movement control in reading. While these models borrow the distinction from classic double-step results, it has never been formally tested within the domains to which the models apply. In the current study, a classic approach to studying double-step stimuli is adapted to a naturalistic scene viewing context.

In summary, the scene-based double step experiment has several concrete aims. Firstly, the assumption that is inherent in several influential models both in scene viewing and in reading is that delays in the latency of fixations can be

partially accounted for by the time required to cancel and reinitiate a saccade program to a novel stimulus. These models often cite basic research into saccadic programming although little work has been done to verify these findings for the more naturalistic case. Therefore, our study investigates saccade programming within scene viewing by adapting a classic double-step paradigm to the scene viewing context. Secondly, by including a classic version of the double-step paradigm as an experimental condition we are able to directly compare performance across tasks.

Method

Participants were all University of Edinburgh undergraduate students that were paid £7 in compensation for their time. Each participant was presented with 100 trials in each condition (Static vs Scene). In the scene condition 100 unique colour photographs were presented at a resolution of 800x600. Stimuli were presented on a 21-inch CRT monitor and participants were seated at a distance of 67cm from the monitor. Eye-movements were recorded with an SR Research Eyelink 1000 desktop system operating at 1000Hz. Out of the 16 subjects tested, 4 were rejected for poor data quality. Of the remaining 12 participants the mean age was 23 and 10 participants were female and 2 were male.

Experiment Overview

Double-step experiments typically involve having a participant fixate to a location while a stimulus is displayed at a distal location. At varying delays, this target is then shifted to a new location. The participant is instructed to make a fixation to the final location of this double-stepped stimuli. At short delays, the participant is frequently able to interrupt whatever processing may have been made to the first target, and instead program a saccade to the second target location. Important aspects of the saccade motor system can be derived by looking at the time course of the response. In order to investigate double-step performance in a more naturalistic environment we adapted a single experimental condition from a classic double-step experiment (Becker & Jürgens, 1979) to a context in which participants received the double-step stimulus while they were actively viewing a natural scene. Furthermore, a replication of the Static double-step condition was included as a baseline measure.

Static Condition

Participants fixated a cross located in the centre of the screen. The first target step was presented after a variable delay of between 2000-3000 ms. The first target step was presented to either the left, or to the right of the fixation cross. The target step delay (TSD), the duration elapsed between presentation of the first and second targets, was either 50, 100, 150, or 200 ms. The presentation side and delay durations were counterbalanced. Furthermore, in order to ensure that participants did not simply postpone their responses and wait until the presentation of the second target step had appeared before making a response, 20% of trials were single step trials in which only a single target was presented to the first target location. These single steps also provide a baseline for saccade response parameters that can be compared to those on the double-step trials. The first target was always presented at exactly 7° on the horizontal axis and the second target was always presented at 14° along the same axis.

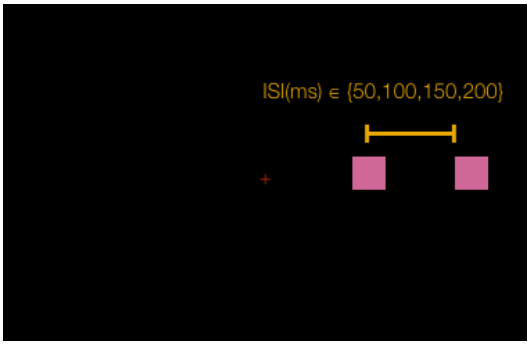

Figure 1. Target steps in the Static condition move in the horizontal plane either to the left or right of fixation cross. The first target step moves to an eccentricity of 7° and the second target step moves in the same direction to an eccentricity of 14°. The second step follows the first at a delay of either 50, 100, 150 or 200 ms. In no-step trials, the first target step to 7° is not followed by a movement to the second target step at 14°.

The instructions to the participant were that they were to "chase the pink box" with their eyes. Their task was to fixate the box as quickly as possible.

After 1000 ms the trial was terminated and a new trial was initiated once the fixation detection procedure had assured that the participants' gaze was directed towards the central fixation cross.

Scene Condition

The instructions to the participant in the scene condition were that they would be required to memorise a scene for a later recall test. However, this recall test was never applied. Furthermore, participants were instructed that they would see pink boxes appear while memorising the image, and that when they see these pink boxes that they should "chase the pink box" with their eyes. The memorisation task was included in order to observe performance in a more naturalistic and cognitively demanding context. All temporal characteristics of the double-stepped stimuli were identical to those of the Static condition. In the scene condition 20% of trials only had a single step at an eccentricity of 7° in order to avoid participants making anticipatory saccades to the second target location at 14°. The first target step was presented once 15 saccades had been made and the scene had been viewed for at least 4 seconds. The first target step was always presented while a fixation was in progress, and this fixation could be at any possible location on the screen. A further difference from the Static double-step condition is that steps were not simply placed on the central horizontal plane as was done in the Static condition. In order to maximize the similarity be-

tween the Static and Scene conditions, while also adapting the study to a naturalistic context, the double-step manipulation trajectory was determined by the line intersecting the current fixation and the last recorded fixation. The first target was presented at 7° along this line in the same direction as the eye-movement plotted from the nth and n-1th saccade. The second target was presented at 14° on the same line. As was done in the Static condition the second target was presented in the same direction as the first (See Figure 2 for details). In circumstances such that projecting the targets along the line of presentation would result in a target being presented off the dimensions of the screen, the presentation procedure was delayed until a fixation occurred such that the presentation of the targets would not occur off screen.

The decision to place the targets along any trajectory intersected by the most recent two fixations was done for two reasons. Firstly, we wanted to control for the effect that angular changes of successive saccades has on resulting fixation durations (Tatler & Vincent, 2008). Furthermore, it is known that saccades in scene viewing are primarily executed along a horizontal axis (Nuthmann & Henderson, 2010). Due to such a bias it was expected that manipulations would primarily be placed along the horizontal axis and this was confirmed with a post-hoc analysis.

Figure 2. In the scene condition targets are presented at 7° and 14°. Unlike the Static condition targets can be placed on any axis within the image. The angle at which the boxes are presented is determined by the location of the current and previous fixations and was presented in the direction of the eye-movement. The delay between target presentations is either 50, 100, 150 or 200 ms. As in the Static condition 20% of trials consisted of only a single step to 7°.

Gaze contingent fixation detection

In order to present targets to participants within the scene condition it was necessary to accurately detect the presence of a fixation with as much temporal precision as possible. The native Eyelink gaze contingent algorithms were used in order to detect the onset of fixation. Once the conditions for presenting the first target had been met, and the Eyelink detected a fixation, the first target was presented to the participant. Delays in the online detection of fixations resulted in the targets being presented after the onset of fixation at a delay (ms) of $\mu = 45.2; \sigma = 19.0$.

Results

Amplitude transition function in the Scene condition

The aim of the first analysis is to provide evidence in support of the hypothesis that saccade cancellation does indeed occur within a more natural scene viewing context than is typically studied with double-step stimuli. In order to construct an ATF, only trials in which both the first and second target steps occurred during a single critical fixation were analysed. Trials were also rejected when the response saccade was not made in a direction consistent with the target steps. Therefore, in the Scene condition 33 trials were removed due to movement of the eyes prior to presentation of the second target, and 6 were removed due to detection of a misdirected saccade. In the Static condition 22 and 2 trials were removed, respectively.

The amplitude transition function for the scene condition was constructed by fitting a four parameter logistic function with a form:

$$y = a + \frac{b-a}{1+e^{c(d-x)}} \tag{1}$$

where a represents a lower bound for the sigmoid, b represents an upper bound, c scales the response to x (Delay) about the midpoint and d is the inflection point of the sigmoid.

Model fitting The data were fit with a nonlinear mixed effects model (NLME) (Pinheiro & Bates, 2000). Firstly, we added fixed effects which allow us to directly estimate the effect of experimental condition (Static vs Scene) on the parameters of the nonlinear response function described in (1). Secondly, random effects were included in the model in order to reduce the effect of unreliable differences between participants due to unbalanced observations and individual variability in task performance.

In the analysis of the Scene condition we fit a model which included the effect of only a single condition on the parameters a, b, c and d. Random effects of participant on the parameters a, b, c and d were also included in the model. For the comparison between the Static vs Scene conditions the model was extended to include a fixed effect of condition on the four model parameters. The R statistical programming language (R Development Core Team, 2008) and the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2013) were used to conduct the analysis.

Effect of Scene on model parameters The parameters of the best fitting model are reported in Table 1. From the scatter the typical ATF evoked by double-step stimuli is observed with a characteristic sigmoidal shape. The horizontal dotted lines indicate the location of target 1 at 7° and target 2 at 14°. Furthermore, the scatter confirms that within the Scene condition lower values of D are typically associated with saccades directed towards the location of the first target step whereas at larger values of D reprogramming occurs and saccades are directed towards the second target location.

Table 1. Mixed effects model parameters

Effect	*M*	*SE*	*t*	*p*
a	6.47	.160	40.48	(< 0.01)
b	12.35	.351	35.23	(< 0.01)
c	0.12	.027	4.33	(< 0.01)
d	113.18	5.46	20.72	(< 0.01)

Summary of the estimated values of the fixed-effects parameters along with their means (M), standard errors (SE), *t* and *p* values, units of the parameters are reported in milliseconds. The parameters *a* and *b* are respectively the lower and upper asymptote of the sigmoid while *c* scales the response about the midpoint and *d* is the inflection point of the sigmoid.

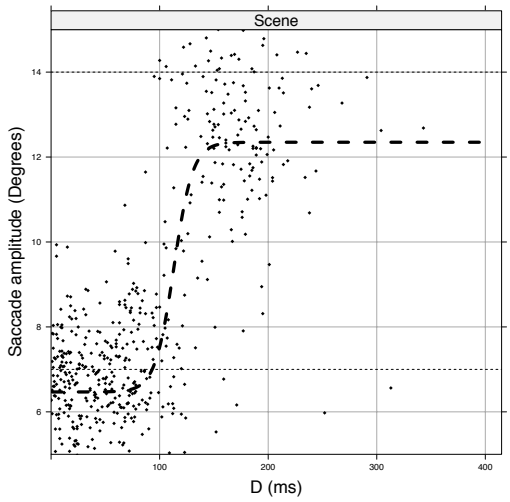

Figure 3. Amplitude transition function constructed from responses in the Scene double-step condition. *D* represents the amount of time elapsed between the onset of the second target stimulus and detection of the first response saccade. Horizontal lines represent the locations of target steps 1 and 2.

The results of the model fits estimated that the lower and upper bounds of the saccadic endpoints were 6.47° (*SE* = .160, $t = 40.48$) and 12.35° (*SE*= 0.35, $t = 35.23$) respectively. While there was undershoot for saccades targeting both the initial and final target steps, the undershoot to the final step was larger (0.53° vs 1.65°).

Cumulative distribution function of saccade latencies

Saccade programming latencies were compared on trials in which there was no target step (no-step) with trials in which there was a target step (50, 100, 150, 200 ms). This analysis aims to investigate whether trials in which a saccade was reprogrammed from the first target to the second target require longer latencies when compared to no-step trials in which no such reprogramming occurs.

In no-step trials saccade latency was calculated as the elapsed time (ms) from the appearance of the 7° target and the first observed response saccade. For trials of all other delays, latencies were analysed for saccades in which the first response saccade occurred after the second target step had appeared and in addition that the first response saccade compensated for the second target step. Compensation was defined such that the saccadic endpoint was within a distance 2° of visual angle from the second target location. Latency was calculated as the elapsed duration between the onset of the first target step and the onset of first response saccade.

A cumulative distribution function (CDF) was fit to compare latencies between delays of different lengths. Latencies from 200 ms TSD trials were excluded as too few compensated saccades were observed. Latencies for compensated saccades are clearly longer than those of saccades elicited by no-step stimuli indicating that in order to incorporate the second target step into the response, increased latencies are required. Furthermore, we observe that as the TSD increases a corresponding increase in latency is also observed.

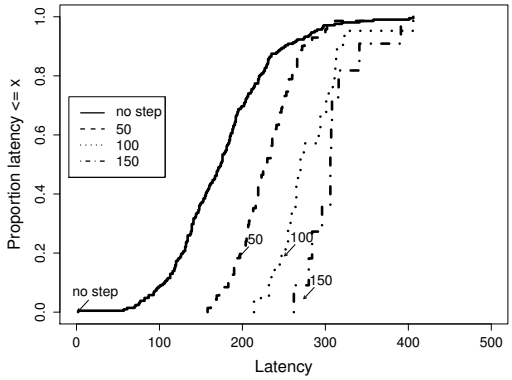

Figure 4. Cumulative distribution function of latencies at target step delays of 50, 100, 150 ms compared to the no-step latencies. The no-step latencies are constructed from latencies on trials in which there was only a step to the first target response.

Comparison of Static vs Scene conditions

While the primary aim was to provide evidence that results from static double-step conditions generalize to a more dynamic scene based context, our dataset also offers an opportunity to directly compare performance differences between the Static and Scene conditions. A description of the model used to fit the Static vs Scene data can be found in the *Model fitting* section of the Scene only analysis.

Comparing the scatter in the Static vs Scene condition (Fig. 5) it is apparent that there is considerably more variability in the data that comprise the ATF in the Scene as compared to Static condition. Due to the more dynamic nature of the Scene task this is to be expected. For instance, in the Static condition participants stay fixated on a central cross while they wait for the target stimuli to appear. It is therefore likely that any anticipatory processes preparing future eye-movements are suppressed. In contrast, during the Scene con-

Table 2. Mixed effects model parameters

Effect	*M*	*SE*	*t*	*p*
a (Intercept)	6.63	0.14	48.62	(< 0.01)
a (Scene)	-0.17	0.16	-1.1	($= 0.27$)
b (Intercept)	13.0	0.11	119.9	(< 0.01)
b (Scene)	-0.74	0.20	-3.82	(< 0.01)
c (Intercept)	0.14	0.02	7.57	(< 0.01)
c (Scene)	-0.05	0.02	-2.18	($= 0.03$)
d (Intercept)	76.14	2.30	32.91	(< 0.01)
d (Scene)	33.86	2.51	13.44	(< 0.01)

Summary of the estimated values of the fixed-effects parameters along with their means (M), standard errors (SE) and t and p values, units of the parameters are reported in milliseconds. The parameters a and b are respectively the lower and upper asymptote of the sigmoid while c scales the response about the midpoint and d is the inflection point. The intercept indicates the estimated parameter in the Static condition, while (Scene) indicates the influence of condition Scene.

dition participants are actively engaged in search, the display and measurement of their double-step response is likely to incorporate processes involved in preparing an eye-movement prior to the presentation of the double-step stimuli.

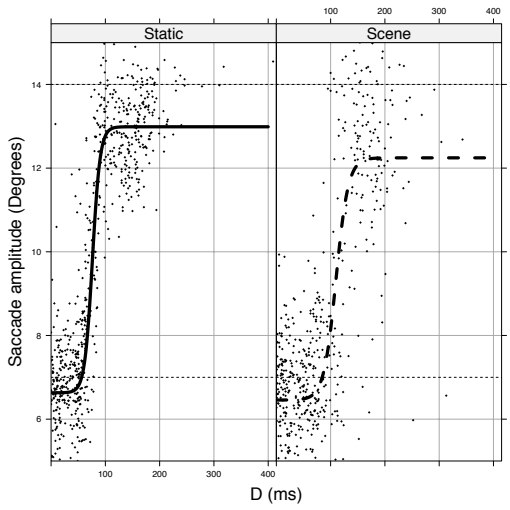

Figure 5. Comparison of the amplitude transition functions in Static vs Scene conditions. *D* represents the amount of time elapsed between the onset of the second target stimulus and detection of the first response saccade. Horizontal lines represent the locations of target steps 1 and 2.

Effect of experimental condition (Scene vs Static) on model parameters Model parameters are summarized in Table 2. The lower and upper bounds of the fitted functions measure the saccadic endpoints of responses to either the first (lower bound) or second (upper bound) targets. No effect of condition was observed on the lower bound ($t = -1.1$, $p = 0.27$) however, there was an effect of condition on the amplitude of responses targeting the final location ($t = -3.82$, $p < 0.01$), indicating greater undershoot in the Scene condition.

A final observation can be made regarding the markedly slower compensation response in the Scene condition. The ATF in the Scene condition appears to be shifted to the right, and this reflects that increasing values of *D* are required to make a response of corresponding amplitude to that of the Static condition. Furthermore, we observe a significant effect of condition on *d* (inflection point) ($t = 13.44$, $p < 0.01$) supporting the observation that responses of comparable amplitude require longer values of *D* in the Scene condition as compared to the Static condition.

Ludwig et al. (2007) have referred to saccadic dead time (SDT) as the last moment at which a new stimulus can modify a saccade program currently under preparation. They describe that the SDT may be extracted from the ATF by estimating the point at which the compensation function begins to incorporate the location of the second target step. The SDT was extracted from our ATF by deriving the point on the curve which represents a cumulative increase of 5% from the lower asymptote. We estimated this point in the Static condition as 55 ms and in the Scene condition as 77 ms.

Discussion

The research question that this paper addresses is whether established results utilizing double-step stimuli to explore saccade programming can be extended to scene viewing. It has been argued that the ATF constructed from double-step responses provides evidence for a distinction between a labile and nonlabile stage of saccadic programming. A target stimulus is only able to modify the current goals of a saccade while it is in the labile stage of programming and can no longer have an influence once the program becomes nonlabile. The sigmoidal shape of the ATF (Fig. 3) reveals that when the target stimulus is presented shortly before the saccade (low values of *D*) that processing had reached the nonlabile threshold and therefore had no influence on the resulting saccade. When the second target is presented in earlier stages of saccade preparation (higher values of *D*) we see saccades that compensate for the second target location due to programming still being within the labile stage. These results have been previously established in double-step studies utilizing static conditions (Becker & Jürgens, 1979; Ludwig et al., 2007), and our study provides evidence for an analogous process occurring within a condition more akin to naturalistic scene viewing.

Cancellation has also been suggested as a causal mechanism for the systematic increase in observed fixation durations. For instance, in the CRISP model, saccade cancellation accounts for the increase in fixation durations that is observed directly following a delay of stimulus onset (see Figure 7, Nuthmann et al., 2010 for details). The CDF (Fig. 4) illustrates that latencies are increased in trials in which a reprogramming of a saccade is likely to have occurred. However, caution must be taken when assigning a causal interpre-

tation to the role of saccade cancellation in observed latency increases. The TSD trials analysed are specifically those for which a saccade was not executed prior to the appearance of the second saccade target. Therefore, we expect to see a complementary increase in latency alongside increases in TSD. One possibility is that increased latencies are observed specifically because compensated saccades are those in which the programming to the first target progressed slowly enough to wait out and incorporate the appearance of the second target. While this analysis does provide confirmatory evidence that saccade cancellation is consistent with increased fixation durations, it does not necessarily shed light on the causal connection between cancellation and increased latencies.

The comparison between the Static and Scene condition also indicate the presence of several notable differences. Ludwig et al. (2007) analyse a concept termed saccadic dead time (SDT) which corresponds closely to the concept of a nonlabile stage of programming. The SDT corresponds to the last point in time at which a saccadic eye-movement may be modified. We estimated SDT as 55 ms in the Static condition and 77 ms in the Scene condition. Differences in the SDT across experimental conditions have been observed in prior work (Ludwig et al., 2007). An important implication of this result with regard to models of gaze control in naturalistic scene viewing is that it provides an empirical bound for the duration of the nonlabile stage. In CRISP for instance, a mean duration of 40 ms was assigned to the duration of the nonlabile stage. This value was determined from classic double-step results conducted under static conditions. This value is also roughly consistent with the duration of the nonlabile stage estimated in our own Static condition (55 ms) but represents an underestimate when compared to the nonlabile duration in the Scene condition. As CRISP is a model of fixation durations in scene viewing it is likely that the estimated mean duration of the nonlabile stage in our Scene condition represents an improvement over the corresponding Static estimate.

It should be noted that the comparison reported here may still reflect important differences not solely attributable to the influence of scene processing. For instance, the Scene but not Static condition double-step targets were presented on any axis. Future work may consider including a task in which performance in our Scene condition is compared directly to a similar task but one in which the scene is replaced by a noise filtered image.

Further comment is warranted on the applicability of the data reported here to models such as CRISP that claim a causal interpretation for saccade cancellation in observed fixation duration delays. In CRISP, when a saccade program is within the labile stage of programming a cancellation signal may interrupt the current program. The time required to reinitiate a new saccade program results in a delay to the current fixation duration. In an alternative formulation aimed at explaining saccade latencies within a Static double-step context, Camalier et al. (2007) suggest that cancellation occurs due to a race between a process initiated to execute a saccade (GO) and process initiated to cancel that saccade (STOP). While there is some similarity in the assumptions between these two models the race model does not insist on a nonlabile/labile dichotomy. Rather, the race model accounts for double-step performance with reference to the timing of the GO and STOP accumulation processes. In order to further explore the role of saccade cancellation it may be of interest to directly compare the predictions of the saccade programming mechanisms in the CRISP model with those of the race model described in Camalier et al. (2007).

References

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967-983.

Camalier, C., Gotler, A., Murthy, A., Thompson, K., Logan, G., Palmeri, T., et al. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*(16), 2187-2211.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777-813.

Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, *97*(1), 795-805.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 1-19.

Nuthmann, A., Smith, T., Engbert, R., & Henderson, J. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382-405.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2013). nlme: Linear and nonlinear mixed effects models [Computer software manual]. (R package version 3.1-108)

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125-157.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18.

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, *52*(6), 932-941.

B

APPENDIX B

Mechanisms of saccadic decision making while encoding naturalistic scenes

R. Calen Walshe

Psychology Department, University of Edinburgh, Edinburgh, UK

Antje Nuthmann

Psychology Department, University of Edinburgh, Edinburgh, UK

Saccadic eye movements are the primary vehicle by which human gaze is brought in alignment with vital visual information present in naturalistic scenes. Although numerous studies using the double-step paradigm have demonstrated that saccade preparation is subject to modification under certain conditions, this has yet to be studied directly within a naturalistic scene-viewing context. To reveal characteristic properties of saccade programming during naturalistic scene viewing, we contrasted behavior across three conditions. In the Static condition of the main experiment, double-step targets were presented following a period of stable fixation on a central cross. In a Scene condition, targets were presented while participants actively explored a naturalistic scene. During a Noise condition, targets were presented during active exploration of a 1/*f* noise-filtered scene. In Experiment 2, we measure saccadic responses in three Static conditions (Uniform, Scene, and Noise) in which the backgrounds are the same as Experiment 1 but scene exploration is no longer permitted. We find that the mechanisms underlying saccade modification generalize to both dynamic conditions. However, we show that a property of saccade programming known as the saccadic dead time (SDT), the interval prior to saccade onset during which a saccade may not be canceled or modified, is lower in the Static task than it is in the dynamic tasks. We also find a trend toward longer SDT in the Scene as compared with Noise conditions. We discuss the implication of these results for computational models of scene viewing, reading, and visual search tasks.

Introduction

To access information-rich regions of the visual field, the visual system engages in eye-movement behaviors known as fixations and saccades. The coordination of such movements involves a complex array of motor control mechanisms operating on distinct spatio-temporal scales. Fixations are defined as the state at which the eye remains in a relatively stable position on some aspect of the visual environment. In scene perception, it is known that the duration of fixations are dependent on such factors as the type of task that people are engaging in (Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011; Nuthmann, Smith, Engbert, & Henderson, 2010; Smith & Mital, 2013), the relevance of the fixated material to the task goals (Land & Hayhoe, 2001), and the lower-level stimulus properties such as the luminance (Henderson, Nuthmann, & Luke, 2013; Walshe & Nuthmann, 2014) and color (Ho-Phuoc, Guyader, Landragin, & Guérin-Dugué, 2012) of the scene. Saccadic eye movements are the primary mechanism by which the eye is brought into spatial alignment with scene content that is to be inspected in high-resolution foveal vision. Where observers fixate is influenced by both mid-level and higher-level stimulus properties (Nuthmann & Einhäuser, 2015). However, it is also known that the effect of image features on fixation selection in scenes may be overridden by task demands (Einhäuser, Rutishauser, & Koch, 2008). Like fixation durations, properties of saccades are also known to vary depending on the task. For instance, participants generate larger saccade amplitudes when searching for an item than when memorizing a scene for later recall (Mills et al., 2011).

Theories of eye-movement control are primarily interested in elucidating the underlying, hidden mechanisms that generate behaviors such as saccades and fixations. The question of what event during stimulus processing results in the initiation of a saccade program to shift fixation away from the currently fixated location is a matter of considerable debate. This question has been most directly addressed in theories of eye-movement control while engaging in reading

Citation: Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision*, *15*(5):21, 1–19, http://www.journalofvision.org/content/15/5/21, doi:10.1167/15.5.21.

doi: 10.1167/15.5.21 Received July 31, 2014; published April 29, 2015 ISSN 1534-7362

behavior in which two competing views have been suggested (Reingold, Reichle, Glaholt, & Sheridan, 2012, for review). The first view suggests that a saccade program is triggered only once the currently viewed stimulus has been processed to a sufficient degree (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2012). Variants of this view are referred to as *cognitive trigger theories*, as saccadic eye movements are generated only when cognition-related processing has been achieved to a sufficient depth. In contrast to the triggering mechanisms just described are those that suggest that the variability in the termination of a fixation is a result of difficulties in lexical processing that interfere with saccade initiation processes. In such models, the decision to initiate a saccade is achieved by an autonomous random timer, and the duration of this timing process may be modulated by the difficulties encountered during stimulus processing (Engbert, Nuthmann, Richter, & Kliegl, 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Therefore, moment-to-moment difficulties in processing result in increased random timing intervals and, consequently, longer fixation durations. Such principles of saccade generation have been adapted to explain fixation duration distributions within complex scene-viewing tasks. The CRISP model (Nuthmann et al., 2010) incorporates a random-walk timer that generates signals to begin the programming of a saccade. Importantly, the CRISP model also allows such saccade programs to be canceled in the case that programming has not proceeded to a sufficiently advanced stage. The CRISP model has been demonstrated to provide good fits to empirical data under a number of experimental conditions (Nuthmann et al., 2010; Nuthmann & Henderson, 2012).

A powerful tool that has been used to reveal the empirical properties of saccade programming timelines is known as the double-step paradigm (Westheimer, 1954; Becker & Jürgens, 1979). Classic variations of the double-step paradigm involve presenting participants with two targets along a horizontal axis with a varying interstimulus interval separating the two targets. For instance, in a seminal study, Becker and Jürgens (1979) had a condition in which a first target was presented at 15° to the left or right of fixation with a second target presented at 30° in the same direction at delays of 50, 100, 150, and 200 ms. The participants' task was to fixate the target as quickly as possible, which thereby placed a saccade program that had been initiated to target the first location in competition with a saccade program targeting the second location. The critical behavioral measure in this study was the amplitude of the saccade that was observed in response to the presentation of the double-step stimulus. Responses observed to fall spatially close to the location of the initial target step permitted the inference that information regarding the updated target location was not incorporated into the programming of the saccadic response. Likewise, the degree to which saccade amplitudes deviated from the initial location and tended to land close to the second target provided a measurement of the degree to which the second target had influenced the resulting saccade. The authors observed an interesting temporal dependency between the amplitude of the saccadic response and the presentation of the second target. If the response saccade occurred in close temporal proximity to the appearance of the second target, then the response saccade tended to fall close to the initial target position. As the temporal interval between the presentation of the second target and the response increased, responses gradually tended toward the final target position. These results revealed that saccades may be reprogrammed when updated position information is available and that the mechanisms underlying this behavior are dependent on the temporal relationship between the updated target position and the onset of the response saccade. This finding has since been replicated by many authors (Aslin & Shea, 1987; Camalier et al., 2007; Findlay & Harris, 1984; Ludwig, Mildinhall, & Gilchrist, 2007; Ray, Schall & Murthy, 2004).

A question that follows directly from such results is at what point in time prior to the onset of a saccade does updated target position information cease to have an influence on the impending saccade? Such a point of no return is often referred to in the literature as saccadic dead time (SDT). Findlay and Harris (1984) analyzed double-step data from a replication of Becker and Jürgens (1979) and found that saccades begin to incorporate targets that are displaced 80 ms prior to saccade onset. More recently, Ludwig et al. (2007) used a double-step task to investigate whether estimated values for SDT show evidence for sensitivity to task effects. In their study, the angle of displacement between the first and second double-step target was manipulated. They found that SDT increased as the angle between the initial and final target locations increased. These results provide empirical support for the claim that SDT values are sensitive to characteristics of the task environment.

The concept of a point of no return has been suggested as an explanation for a number of empirical effects observed within the scene-viewing literature. In a procedure known as the stimulus onset delay (SOD) paradigm (Henderson & Pierce, 2008; Shioiri, 1993) participants viewed naturalistic scene stimuli while engaged in an encoding task. During a critical fixation, the scene was removed and replaced by a mask, thereby removing scene content from further processing. Within the same fixation, the scene was then restored to view at varying delays. Results from the SOD paradigm showed a bimodal distribution of fixation durations

such that one population was independent of the length of the delay whereas the other population increased in proportion to the length of the delay. In the context of the CRISP model, Nuthmann et al. (2010) suggested that such bimodality arises due to two factors. First, when the scene disappears from view, the rate at which saccade programs are generated slows down to reflect the lack of incoming visual information. Second, if a saccade program is currently within the labile stage of saccade programming (i.e., has not passed the point of no return), then the current saccade program is subject to stochastic cancellation. Together, these processes combine to yield delays in the programming of saccades and therefore longer fixation durations. Wu et al. (2013) also demonstrated an influence of a point of no return on saccade programming in a study in which a scene was swapped with a novel scene while participants were engaged in a viewing task. In the saccade that immediately followed the change, it was found that participants systematically programmed saccades to target the center of the screen. However, such a center scene bias occurred primarily for saccades that were initiated at relatively long durations after the change. For saccades that were initiated shortly after the display change, there was no evidence for such a bias. Similar to the SOD paradigm, such a result is well accounted for by the fact that in those saccades that did not target the center, the scene change occurred when saccade programming had already passed the point of no return and could therefore no longer influence saccade programming.

The principles derived from such investigations with double-step stimuli have provided the basis for the implementation of eye-movement control models in scene viewing as well as in reading. In the E-Z Reader model (Reichle et al., 1998), the authors introduced the concepts of a labile and a nonlabile stage of saccade programming. This distinction implies a multistage saccade programming architecture whereby saccade programming that is within the labile stage is subject to cancellation. However, once programming has progressed beyond the labile stage into the nonlabile stage, cancellation is no longer possible. Such multistage saccade programming assumptions have since been incorporated into a number of models that explain oculomotor control under a variety of task conditions such as scene viewing (Nuthmann et al., 2010), reading (Engbert et al., 2005; Pollatsek, Reichle, & Rayner, 2006; Reichle et al., 1998), and visual search (Trukenbrod & Engbert, 2014).

One notable difference between these models is the assumptions that are made regarding the duration of the nonlabile stage. In both the CRISP model (Nuthmann et al., 2010) and the E-Z Reader model (Reichle et al., 1998), it is assumed that the nonlabile stage adopts a fixed duration and is not dependent on stimulus characteristics. On the other hand, as of SWIFT-II (Engbert et al., 2005), the model makes the explicit assumption that the duration of the nonlabile stage may vary depending on the amplitude of the planned saccade. Therefore, the SWIFT model explicitly incorporates systematic variability in SDT while the CRISP model and E-Z Reader do not. It is important to note that all models just introduced do incorporate *random* (i.e., unsystematic) variability in the duration of the nonlabile stage as the duration of a given nonlabile stage is drawn from a statistical distribution. Although Ludwig et al. (2007) provides some empirical support for stimulus-dependent differences in the duration of the nonlabile stage, there is currently no empirical research demonstrating task-specific differences within reading or scene-viewing contexts. Furthermore, upon inspection of the model parameters that are used to describe the duration of the nonlabile stage, it becomes evident that there is very little consistency between (and within) models regarding the duration of this stage. We return to this issue in the Discussion.

The aim of the current study is to investigate saccade programming during naturalistic scene perception by embedding a double-step task within scene-viewing contexts. Two primary questions are addressed with these experiments. The first addresses an empirical gap in the scene-viewing literature by testing whether saccade cancellation operates in an analogous manner within scene viewing as it does within classic double-step investigations. The assumption that findings from low-level tasks generalize to high-level task contexts has often been made by models of oculomotor control (Engbert et al., 2005; Nuthmann et al., 2010; Reichle et al., 1998; Trukenbrod & Engbert, 2014). However, it is an open question that has remained largely unaddressed (but see Walshe & Nuthmann, 2013). Our second question asks whether SDT values vary depending on experimental context. By deriving such an estimate, we also explore how such a value may provide an indirect measure of the duration of the nonlabile stage by taking into account the delays in transmission of information between the retina and cortical regions responsible for saccadic decisions.

In Experiment 1, we address these questions by comparing double-step performance across three conditions. The first condition (Static) replicates a classic version of the double-step procedure (Becker & Jürgens, 1979). In a second condition (Scene), a scene-viewing double-step task is conducted by presenting the double-step targets during active exploration of scene content. The third condition (Noise) replicates the experimental design of the scene-viewing double-step task but instead replaces the scene with a phase noise–transformed background stimulus. Such a transformation is achieved by applying noise to the phase spectrum but leaves the amplitude spectrum intact

(Einhäuser et al., 2006). Applying a transformation in this manner removes object and other higher-order scene statistics yet retains the $1/f$ characteristics of the amplitude spectra that is typical of naturalistic scenes (Einhäuser et al., 2006). In a follow-up experiment, we test how SDT is influenced by scene background independently of dynamic movement context. In Experiment 1, observed differences in SDT between the Static condition and the dynamic movement conditions (Scene, Noise) could be attributed to either the background content or differences in the movement context. Experiment 2 provides a test of these two hypotheses by comparing the three backgrounds used in Experiment 1 in static movement contexts only.

To make comparative inferences about saccade programming across conditions, we use a nonlinear mixed-effects (*nlme*) regression framework that improves on previous methods of analyzing double-step performance. Population-level parameters provided a method to compare performance across conditions, whereas individual parameter estimates were extracted from the fitted model and were used to provide a by-participant measure of SDT in the three conditions.

Experiment 1

Methods

Participants

Two men and 13 women (mean age = 23 years) recruited from the University of Edinburgh student population participated in the study. Participants completed all experimental conditions in one session, which lasted approximately 1.5 h. Each participant was paid £7 per hour of participation in compensation for their time. The study conformed to the tenets of the Declaration of Helsinki, and written consent was supplied by the participants prior to the experiment.

Apparatus

Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz, and the monitor screen was at a distance of 67 cm from the participant. During stimulus presentation, participants' eye movements were recorded using an SR Research EyeLink 1000 Desktop mount system. It was equipped with the 2000-Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz for each eye with an average spatial accuracy of 0.25°–0.5° of visual angle. Viewing was binocular, and both eyes were tracked. Only the position of the right eye was used in the analysis. A chin rest was used to achieve stability of a participants' head position relative to the screen. The experiment was implemented in MATLAB 2009b using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007), which incorporates the EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002). The software allowed precise control over the timing of display changes. To detect fixations online, we implemented a nine-sample online velocity estimation algorithm in MATLAB that aimed to mimic Data Viewer's offline velocity estimation procedure (SR Research Ltd., 2006). Fixations were detected offline using SR Research Data Viewer to parse the gaze samples into sequences of fixations and saccades.

Stimuli

In all conditions, the targets consisted of isoluminant 1.5° square boxes presented in the color pink (CIE_{Lab} L = 65.48, a = 61.84, b = −26.03). In the Static task, the background was uniformly black. When participants were required to fixate on a central cross, it was presented in red (CIE_{Lab} L = 53.23, a = 80.42, b = 66.96). In the Scene task, participants viewed images of 200 naturalistic scenes, in addition to four practice scenes. Each scene had a resolution of 800 × 600 pixels and was presented in full color. Scenes were collected from online databases such as Flickr and Google images. They were selected to include a variety of categories such as indoor and outdoor as well as urban and nature scenes. Each scene was viewed by the participant only once over the duration of the experiment. At a viewing distance of 67 cm, the scenes subtended 33° × 25°. The stimuli in the Noise task consisted of 200 background images that were constructed by applying a phase-noise filtering procedure to the images used in the Scene task. The application of the phase noise–filtering process allowed removal of higher-order scene statistics such as those used to determine edges and contours while leaving the amplitude spectrum unmodified. Phase noise stimuli were created by transforming the original scene images into Fourier space, where additive noise drawn from a uniform distribution was added to the phase spectrum (Einhäuser et al., 2006). An inverse Fourier transformation was then applied to the images to convert them back to image space.

Procedure

Three double-step tasks were conducted to compare characteristics of saccade programming in static and dynamic gaze conditions. The order of the conditions was counterbalanced across participants. In each of the three tasks, participants completed 200 trials that were classified as either double-step (40%) or single-step (60%) trials. Single-step trials were included to ensure that

participants could not make accurate anticipatory saccades to the final resting location of the double-step target. On single-step trials, the center of the 1.5° pink box was presented at an eccentricity of 7° from the point at which the participant was currently fixating. The order of single-step and double-step trials was randomized.

On double-step trials, two targets were presented at distances of 7° and 14° of visual angle from the fixated location. The first target is referred to as the *initial target* and the second target as the *final target*. Both targets were always presented in the same direction relative to the fixation location. Therefore, because the initial target was presented at 7° from the fixation location, the final target was always presented at a further 7° in the same direction. The color and luminance of the stimuli in the double-step condition were identical to that of the target in the single-step condition. The timing of the initial target step was identical to that of the timing of the single-step stimuli described above. The initial target step was presented for varying amounts of time prior to the onset of the final target step. The interval defining the amount of time elapsed between the onset of the first target step and the onset of the second target step is referred to as the *target step delay* (TSD). The final target was presented simultaneously with the disappearance of the first target. Therefore, the subjective impression of this procedure is that the first target step *jumps* to the second target location.

The TSD for a trial was defined in an adaptive manner such that TSD varied depending on the amplitude of the response on the previous double-step trial (Camalier et al., 2007). A *compensated saccade* refers to a saccade that was programmed to go directly to the final target location, whereas a *noncompensated saccade* is one in which the saccade was programmed to the initial target location. From previous investigations, it is known that short TSDs tend to result in final target response saccades whereas longer TSDs tend to result in initial target responses (Becker & Jürgens, 1979; Camalier et al., 2007). Following a double-step trial in which a compensated saccade was produced, TSD was increased and TSD was decreased following noncompensated saccades. The adaptive increment was defined as 50 ms; a lower bound on TSD was set at 50 ms. The purpose of this adaptive procedure was to balance the number of compensated and noncompensated response saccades. Compensated saccades were detected online and were identified when a saccade was made within 2° of the final target. The presentation of the stimulus was synced with the vertical retrace of the monitor. Full presentation of the stimulus was therefore delayed from the defined TSD by up to 7.14 ms. The trial was terminated 1,000 ms following the first saccade made in response to the presented targets.

Static task

In the Static condition, participants initiated a trial by fixating on a cross presented at the center of the screen. If the eye tracker could not reliably detect fixation on the cross, a recalibration routine was initiated. Following a randomly timed delay of 2000 to 3000 ms, targets were placed on the same horizontal axis as the fixation cross and were presented either to the left or to the right. On double-step trials, the final target was always presented on the same side of the cross as the initial target. Targets were presented to the left or right with equal frequency, and the side was randomly selected for a particular trial. Participants were instructed to fixate the box as quickly and accurately as possible. The design in the Static condition is visualized in Figure 1.

Dynamic tasks

In the dynamic tasks (Scene, Noise), participants initiated a trial by fixating on a central cross. Once the cross had been fixated, the image was revealed and the participant was free to make unconstrained fixations on the screen. In both conditions, the task was to encode the presented image for later recall. Participants were instructed that the recall phase would commence once all the stimuli had been presented. Once the participant had made 10 saccades on the current trial, a critical fixation was identified and the targets were presented. The targets were presented immediately upon detection of the onset of a critical fixation. Fixations were detected online using a custom nine-point velocity estimation algorithm (see the Apparatus section). A primary difference between the static and dynamic tasks is the trajectory at which the targets were presented. In contrast to the Static task, in the dynamic tasks the targets could be presented along any axis and the location of the targets was determined by the position of the current and preceding fixation locations. The targets were presented along an axis that was derived by intersecting the coordinates of the current fixation with the previous fixation. The initial target was therefore placed at a distance of 7° from the currently fixated location in the same direction as the most recent saccade (see Figure 1 for details). The decision to place the targets along such a trajectory was done to control for systematic viewing biases that are known to exist during the exploration of naturalistic scenes. First, it is known that during scene-viewing tasks, saccades are most frequently programmed in the horizontal direction (Foulsham, Kingstone, & Underwood, 2008; Nuthmann & Henderson, 2010; Tatler & Vincent, 2008). We confirmed that such a systematic bias exists in our dynamic conditions by visual inspection of radial histograms showing saccade angles relative to the horizon (see Figure 2). To account for

Figure 1. Experiment 1 design. (A) An example stimulus from the 1/*f* Noise condition is presented. The procedure for the double-step condition in the dynamic tasks is illustrated in (B). The experiment begins with the participant fixating on a central cross. The participant then views a scene (or 1/*f*) stimuli for later recall. On the 10th fixation, the first target is presented immediately upon detection of fixation onset, followed by the second target at intervals of 50, 100, 150, or 200 ms. An adaptive staircase method was used to define the interval on a specific trial. The dashed blue circle represents an example fixation location before target presentation, and the red circle represents an example fixation location after a target response is generated. A *compensated saccade* occurs when a saccadic response is generated directly to the final position of the target (top left of panel B). A *noncompensated saccade* occurs when a response is erroneously generated to the first target position (top right of panel B). (C) The single-step condition is illustrated, in which only a single target is presented. (D) The procedure for a double-step trial in the Static condition. Participants fixate on a red cross at the center of the screen. Targets then appear at intervals of 2000 to 3000 ms. Trials can either be single-step or double-step trials, and the timing is the same as in panels B and C.

such a horizontal bias, we implemented a static double-step task in which targets were presented exclusively on the horizon. This was done to maximize the match between target trajectories in the static and dynamic conditions. Second, we wanted to control for the effect of *saccadic momentum*. Saccadic momentum refers to the tendency for fixation durations to increase as the angle between the preceding and subsequent saccade direction increases (Wilming, Harst, Schmidt, & König, 2013). For example, saccades that are programmed perpendicular to the direction of the preceding saccade result in longer fixation durations. Therefore, by placing both targets along the same trajectory, any influence on the timing of the saccade due to saccadic momentum is minimized. The design in the dynamic tasks is visualized in Figure 1.

Results

Prior to analysis, the data were preprocessed to exclude cases that did not conform to sufficient data quality standards. Each participant viewed 200 trials in each condition, 40% (80) double-step trials and 60% (120) single-step trials. In double-step trials, we excluded any trial in which the response saccade was initiated prior to the time at which both targets had been presented. This left an average of 61 double-step trials in the Static condition, 74 in the Scene condition, and 72 in the Noise condition. Saccades that were clearly too short (<2°) to be considered as responses to either of the targets were excluded. Furthermore, if a blink occurred immediately before or immediately after the fixation in which the targets were presented, this

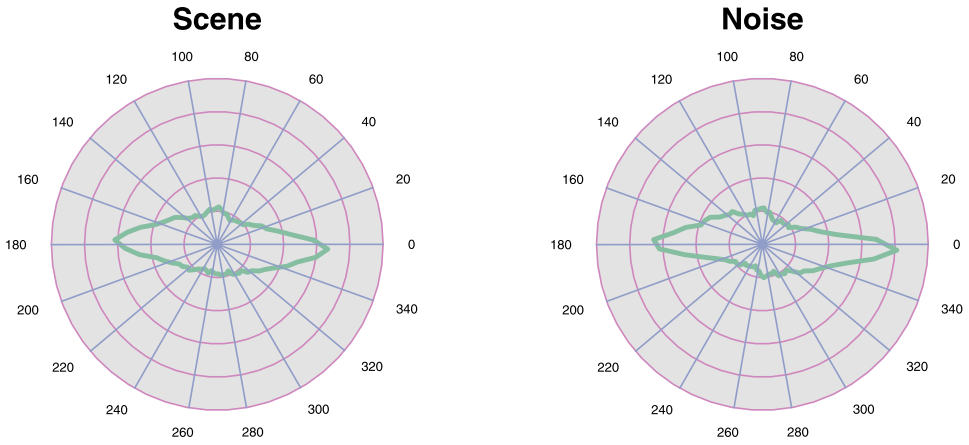

Figure 2. Distribution of saccade direction in Experiment 1. In both the Scene and Noise conditions, there is a clear preference to program saccades along the horizontal axis. Angles of 0° and 180° indicate saccades programmed along the horizon. Densities were calculated from a bin size of 7.2°.

trial was excluded. In the Static task, if the targets were presented when the fixation deviated by more than 2° from the fixation cross, this trial was excluded. An additional criterion was defined such that if a saccade was programmed more than 45° away from the direction that the targets were presented in, these saccades were considered not to be programmed in response to detection of the targets and were excluded from the analysis. After all exclusion criteria had been applied, there remained an average of 56 double-step trials in the Static condition, 67 trials in the Scene condition, and 68 trials in the Noise condition.

Modeling the amplitude transition function

The amplitude transition function (ATF) relates the resulting saccadic response amplitude to a quantity referred to as delay (D). D measures the time elapsed between the onset of the second target step and the onset of the response saccade (Becker & Jürgens, 1979). The amplitude of the saccade provides information about whether a saccade was successfully reprogrammed or not. In the present experiment, amplitudes of approximately 7° indicate saccades programmed to the initial target, whereas amplitudes greater than 7° can provide evidence for the saccade being executed toward the final target location.

Values of D provide a measurement of how much time has elapsed between the onset of the second target and the onset of the saccadic response. Therefore, D measures the amount of time available to reprogram a saccade to the new target location. If the value of D is high, this implies that the second target was available for a relatively long period prior to the onset of the saccade. When the value of D is low, the saccade was executed very shortly after the presentation of the second target. By combining D with the amplitude of the response, thereby constructing an ATF, it is possible to ask the following question: What is the minimal amount of time prior to the onset of the saccade that the second target must be presented to have an influence on the resulting saccade?

Previous research has demonstrated that the ATF in double-step tasks may be well described by a curve that closely resembles the logistic function (Becker & Jürgens, 1979; Ludwig et al., 2007). This was confirmed by graphical analysis of the data in the present experiments as well as ATFs previously reported in the literature. The following four-parameter logistic function was used to model response amplitude as a function of D in the three experimental conditions:

$$f(x) = \alpha + \frac{\beta - \alpha}{1 + e^{\gamma(\delta - x)}}, \quad (1)$$

where α represents the lower bound for the logistic function, β represents the upper bound, γ is a scaling parameter, and δ defines the inflection point.

As a novel approach, ATFs were estimated with a nonlinear mixed-effects regression framework using the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2014) implemented in the R software for statistical computing (R Development Core Team, 2012). Using this approach, we model the variability in the estimation of parameters contributed by both individual participants and experimental condition. Specifically, a nonlinear mixed-effects model provides a method of simultaneously estimating the fixed (population-level) parameters and the random (individual-level) parameters (Pinheiro & Bates, 2000). A benefit of estimates derived from such a framework is that participant-level parameter estimates are weighted by their corresponding population parameters, providing a measure of protection against overfitting the ATFs to

Fixed effects				Random effects	
Parameter	Estimate	SE	*t* value	Parameter	σ
Lower bound (α)					
Static (intercept)	6.15	0.19	31.56	Static	0.59
Noise	−0.07	0.18	−0.39	Noise	0.42
Scene	0.164	0.173	0.946	Scene	0.20
Upper bound (β)					
Static (intercept)	13.05	0.15	85.62	Static	0.47
Noise	−1.18	0.21	−5.50	Noise	0.58
Scene	−0.68	0.24	−2.81	Scene	0.70
Inflection (δ)					
Static (Intercept)	90.05	2.30	39.10	Static	6.77
Noise	16.49	2.31	7.14	Noise	4.15
Scene	33.10	2.10	11.04	Scene	6.80
Scale (γ)					
Static (Intercept)	0.09	0.01	8.39	Static	0.03
Noise	0.07	0.01	1.94	Noise	0.01
Scene	−0.02	0.02	−1.26	Scene	0.05
Error term				ε	1.64

Table 1. ATF regression model (Experiment 1). *Notes:* Estimated parameters for the four-parameter logistic regression model. Means, standard errors, and *t* values of fixed effects; standard deviations of the random effects.

individual-level data. By-participant random effects (intercept and slope) were included for all four parameters of the logistic function, thereby allowing all parameters of the model to vary in a by-participant manner. The random effect covariance matrix was assumed to follow a block diagonal structure. That is, correlations between random intercepts and slopes were permitted only when grouped within the logistic parameters α, β, γ, and δ and were assumed to be 0 otherwise. For example, correlations between the upper-bound intercept and slopes were estimated, but correlations between the upper- and lower-bound random effects were assumed to be 0. Parameters were estimated by maximizing the log-likelihood of the model given the observed responses. The parameters of the fitted model are presented in Table 1.

Saccade modification

The ATF is a direct way to measure the behavioral correlates of mechanisms underlying saccade reprogramming as it measures the amplitude of saccades that are initiated under conditions in which updated target position information is available. Furthermore, the ATF plots the response amplitude as a function of the duration that has elapsed since the presentation of the second target (D). Therefore, the ATF allows the relationship between the time spent processing the second target and the amplitude of the response to be revealed. The ATF allows specific predictions about saccade cancellation processes to be tested. In the case that saccades can never be modified, the ATF predicts a flat function with an intercept close to the location at which the first target is presented. In the present experiment, the predicted intercept would be 7°. In the case in which modification is always possible, a flat function would also be predicted, but in this case, the intercept would be predicted to be near the location at which the second target is presented or 14° in the present experiment. An intermediate hypothesis between these two extremes is that modification of a saccade program is possible and becomes increasingly likely at greater temporal separation between the onset of the second target and the onset of the saccadic response. In this case, a monotonically increasing ATF would be predicted with a lower asymptote located close to the first target location that gradually increases and asymptotes close to the final target location. Inspection of Figure 3 reveals that the shape of the ATF in all three experimental conditions conforms to the intermediate hypothesis. Saccade programs can be modified, and as greater values of D are observed, a saccade targeting the final location becomes increasingly likely.

Furthermore, the regression modeling reveals that, for double-step trials, saccades significantly undershoot the targets and that the degree of undershoot depends on experimental condition. The Static condition was selected as the intercept for each of the logistic parameters that were estimated. Effects were determined to be significantly different from 0 when $|t| > 1.96$ was observed. For the lower bound, the estimated amplitudes for all conditions were less than the distance at which the targets were placed (7°). Task did not significantly influence the lower bound. That is, the estimated change in lower bound from the Static condition (intercept) was not significant for either the Noise or Scene condition. Significant task effects were observed for the estimated upper bound. The upper bound estimate for the Static condition (intercept) was $\hat{\beta} = 13.05°$, $t = 85.62$, which indicates an undershoot to the second target. The decrease in upper bound was greatest in the Noise condition, $\hat{\beta} = -1.18°$, $t = -5.50$, followed by the Scene condition, $\hat{\beta} = -0.68°$, $t = -2.81$. See Table 1 for details.

Saccadic dead time

SDT may be defined as the last point in time at which novel stimulus information may be incorporated by the system responsible for preparing a saccadic response. Thus, SDT may also be described as the point of no return in the preparation of a saccade. Once the point of no return in saccade programming has been

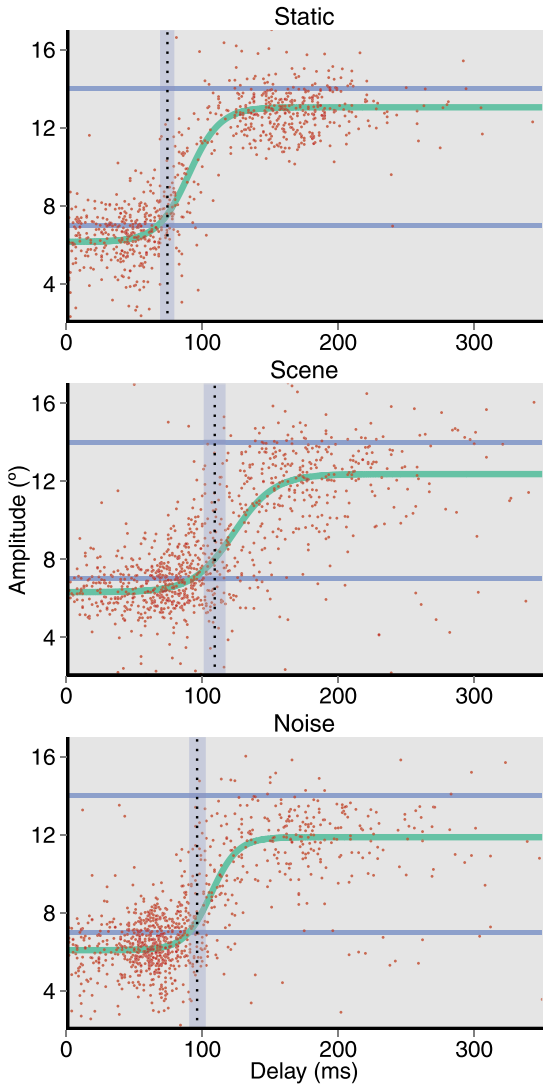

Figure 3. Amplitude Transition Functions for the three experimental conditions in Experiment 1. Points represent the amplitudes of saccades initiated in response to the presentation of double-step targets. On the *x*-axis, the delay (D) represents the amount of time elapsed between the presentation of the second target and the onset of the saccade. The *y*-axis represents the amplitude of the resulting saccade. The green lines represent the best-fitting fixed-effects curves estimated with a nonlinear mixed-effects regression. The horizontal solid blue lines represent the physical distance of the targets in relation to fixation location (7° and 14°). The vertical dotted black lines represent the mean SDT in each condition derived from the by-participant SDT estimates. The light blue bands surrounding the means are bootstrapped confidence intervals.

reached, that saccade may no longer be modified or canceled.

To estimate the SDT from the fitted ATFs, we determined the largest value of D for which the amplitude of the response saccade showed no evidence of incorporating the second target position into the response. To define saccade amplitudes in the double-step condition that are inconsistent with responses programmed to the first target location, we used the distribution of response amplitudes observed in the single-step condition. In the single-step condition, saccades target the first location and are characterized by a distribution of amplitudes that cluster near the first target location. For each participant and each condition, we measured the amplitude that corresponds to the 95th percentile of responses. Such a value provides a cutoff for amplitudes that are rarely associated with responses to the initial target location. This cutoff was used to define an amplitude threshold in the double-step condition such that responses with an amplitude beyond this point were considered to be influenced by the second target step.[1]

Participant	Static	Scene	Noise
1	79	115	94
2	57	127	95
3	67	103	83
4	75	89	89
5	73	105	96
6	83	111	100
7	85	115	106
8	79	112	100
9	69	115	92
10	83	118	114
11	86	134	115
12	70	90	85
13	65	90	85
14	84	122	104
15	65	95	85
Mean	74	109	96

Table 2. Estimated Saccadic Dead Time (ms) in Experiment 1.

As the ATF was estimated with a nonlinear mixed-effects regression with by-participant random intercepts and slopes of experimental condition, it was possible to provide an independent estimate of SDT for each participant in each condition. Specifically, estimates of SDT were computed for individual participants by inspecting the responses predicted by the individual-level data from the mixed-effects regression. Individual-level SDT estimates, along with mean SDT in the three conditions, are provided in Table 2. To statistically validate the differences in observed means, we used a bootstrap procedure to estimate the distribution of mean SDT in each of the three conditions (Efron & Tibshirani, 1993). In summary, the SDT observations recorded in each condition were resampled with replacement 10,000 times, and for each iteration, a sample mean was recorded. Through this method, we constructed a distribution of sample means based on bootstrapped data. These distributions were then used to construct 95% confidence intervals (CIs)

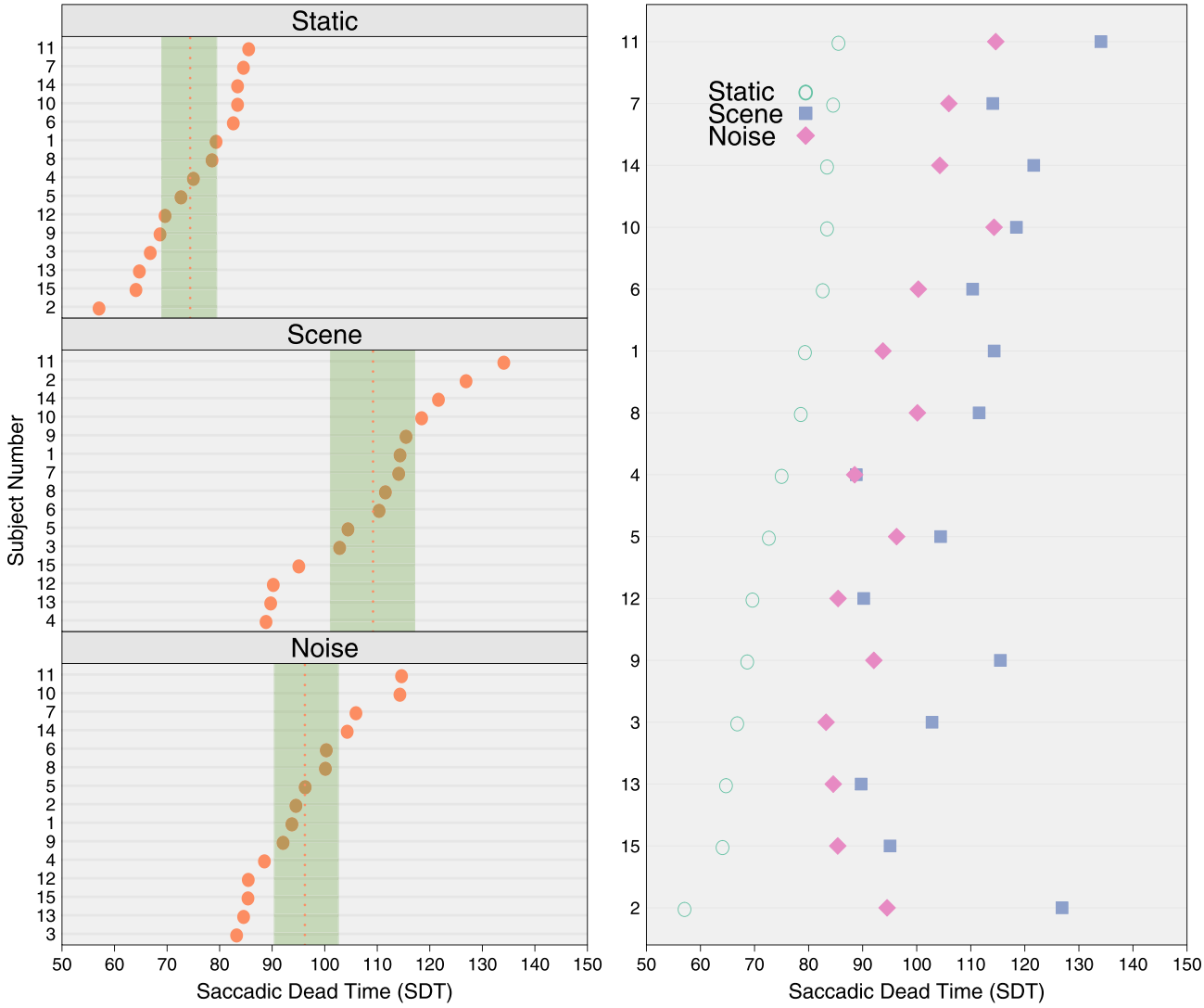

Figure 4. Saccadic Dead Time (SDT) estimates (ms) for the three experimental conditions in Experiment 1. The left plot shows by-participant estimates of SDT. For a given condition, from bottom to top, values are ordered from lowest to highest. The vertical dotted lines represent the mean value of SDT in each condition. The green band surrounding the mean is the bootstrapped confidence intervals of the mean estimate. The right plot shows participants ordered by the magnitude of their estimated SDT value in the Static condition.

around the observed means. We defined a significant difference between means in any two conditions as occurring when the bootstrapped CIs did not overlap. To control for familywise error rates, Bonferroni correction was applied to the CIs. The mean SDTs and CIs were derived from the fitted model. The mean SDT was shortest in the Static condition ($M = 74.35$ ms, $CI_{low} = 69.03$, $CI_{high} = 79.65$), followed by the Noise condition ($M = 96.24$ ms, $CI_{low} = 90.49$, $CI_{high} = 102.69$) and the Scene condition ($M = 109.18$ ms, $CI_{low} = 100.94$, $CI_{high} = 117.27$). Therefore, according to the bootstrapped hypothesis-testing procedure, statistically reliable differences were observed between SDT in the Static versus Scene and Static versus Noise conditions. Furthermore, a trend was observed in the Scene versus Noise conditions such that SDT tended to be longer in the Scene than in the Noise condition. Inspection of the individual measures of SDT (see Figure 4) shows that the SDT for each participant was numerically larger in the Scene than in the Noise condition.

Additional analyses

Additional analyses explored whether the structural differences in the stimulus content in the two dynamic tasks were associated with differences in global eye-movement parameters. To compare mean saccade amplitudes and fixation durations in the Noise and Scene conditions, only those saccades that were not

generated in response to a target step were considered. For saccade amplitudes, no statistically significant difference was found ($p > 0.05$). Saccade amplitudes to single-step targets were also analyzed. They were shortest in the Noise condition ($M = 6.20°$), followed by the Static condition ($M = 6.44°$) and the Scene conditions ($M = 6.51°$). Pairwise t tests with a family-wise error rate of 0.01 were used to statistically validate the mean differences. The difference between Scene and Noise was statistically significant, $t(14) = 3.49$, $p = 0.004$, as was the difference between Static and Noise, $t(14) = 2.80$, $p = 0.01$. The difference between Static and Scene did not reach statistical significance.

Mean fixation durations were longer in the Noise condition ($M = 324$ ms) compared with the Scene condition ($M = 276$ ms), and this difference was significant, $t(14) = 4.28$, $p = 0.0008$. This pattern of results replicates a previous study that compared $1/f$ filtered scenes with naturalistic scenes (Kaspar & König, 2011).

In addition, we analyzed saccadic reaction times to single-step stimuli. Saccadic responses were fastest in the Noise condition ($M = 162$ ms), followed by the Scene ($M = 177$ ms) and Static ($M = 195$ ms) conditions. Saccadic response times were significantly longer in the Scene than in the Noise condition, $t(14) = 2.51$, $p = 0.03$. The difference between the Static and the Noise condition was also significant, $t(14) = 4.19$, $p = 0.001$. The difference between the Static and the Scene and fitted ATFs are shown in Figure 5.

Experiment 2

A second experiment was conducted to compare SDT under conditions in which the structure of the background is varied but in which participants do not explore the scene prior to target onset. In this experiment, participants conducted three versions of the Static task from Experiment 1. In the first condition, the task is conducted on a uniformly black background. We call this condition the Uniform condition and note that it is identical to the Static condition in Experiment 1. In the second condition (Scene condition), the same task is conducted, but the background is replaced by one of the naturalistic scenes used in Experiment 1. In a final condition, the task is conducted with $1/f$ stimuli presented as the background.

As movement is controlled for in this study, any observed difference in SDT can be more directly interpreted to result from differences in the structure of the background. Specifically, we predict that if increases in SDT observed in Experiment 1 are due to the additional structure of the backgrounds in the Scene and Noise conditions, then the structured background conditions in Experiment 2 should reveal an elevated SDT relative to the uniform background.

The experiment was conducted on an additional seven male (including one author, R.C.W.) and five female participants with an average age of 24 years who did not participate in Experiment 1. The three conditions in Experiment 2 (Uniform, Scene, Noise) paralleled in nearly all details the Static condition from Experiment 1. The primary difference was that in two of the conditions, the backgrounds were replaced by scenes (Scene conditions) or phase noise images (Noise condition) that were used in Experiment 1. On each trial, the specific scene or noise image was randomly selected from the set of stimuli used in Experiment 1. Participants were explicitly told that the scene or noise background content was not relevant to the completion of their task.

Results

The approach to the analysis of the results in Experiment 2 was conducted in an analogous way to the Static condition in Experiment 1. The nonlinear mixed-effects model in Experiment 2 included the same random effects structure as the model in Experiment 1. That is, by-participant random intercepts and slopes were included for α, β, γ, and δ. The model estimates are summarized in Table 3.

The estimated SDT in the uniform condition was $M_{Uniform} = 71.39$ ms, $CI_{low} = 65.00$, $CI_{high} = 78.26$. In the Scene condition, SDT was estimated at $M_{Scene} = 80.70$ ms, $CI_{low} = 75.50$, $CI_{high} = 86.11$ and was estimated at $M_{Noise} = 70.83$ ms, $CI_{low} = 65.93$, $CI_{high} = 76.55$ in the Noise condition. As was introduced in the analysis of Experiment 1, lack of overlap in Bonferroni-corrected CIs was used as a criterion for rejecting the null hypothesis. Accordingly, there was no difference in SDT between the Uniform and Noise conditions, as is evidenced by the strongly overlapping CIs (Figure 6). The CI for the Scene condition slightly overlapped with the other two CIs. However, inspection of the pattern of results at the level of individual SDT responses revealed a strong trend for longer SDT in the Scene condition when compared with the uniform and phase conditions. For 10 of the 12 participants, the longest SDT response was observed in the Scene condition (Figure 6; Table 4).

We also analyzed saccade latencies on trials in which only the initial target was presented. Latency was lowest in the uniform condition ($M = 203$ ms), followed by the Noise condition ($M = 208$ ms) and Scene condition ($M = 214$ ms). The differences between Scene versus Noise and Noise versus Uniform did not reach statistical significance, but the difference between Scene

	Fixed effects			Random effects	
Parameter	Estimate	SE	*t* value	Parameter	σ
Lower bound (α)					
Uniform (Intercept)	6.55	0.15	44.15	Uniform	0.42
Noise	0.37	0.13	2.96	Noise	0.10
Scene	0.35	0.12	2.92	Scene	0.02
Upper bound (β)					
Uniform (Intercept)	13.06	0.16	79.48	Uniform	0.52
Noise	−0.08	0.14	−0.57	Noise	0.33
Scene	−0.10	0.13	−0.74	Scene	0.34
Inflection (δ)					
Uniform (Intercept)	86.62	3.79	22.88	Uniform	12.63
Noise	2.71	2.90	0.94	Noise	8.63
Scene	9.96	2.31	4.32	Scene	6.34
Scale (γ)					
Uniform (Intercept)	0.10	0.01	7.83	Uniform	0.04
Noise	−0.01	0.01	−1.04	Noise	0.02
Scene	−0.004	0.02	−0.22	Scene	0.04
Error term					
ε					1.11

Table 3. ATF regression model (Experiment 2). *Notes:* Estimated parameters for the four-parameter logistic regression model in Experiment 2. Means, standard errors, and *t* values of fixed effects; standard deviations of the random effects.

and Uniform was statistically significant, $t(11) = 0.04$, $p < 0.04$.

Saccade amplitudes to single-step targets were also analyzed. They were shortest in the uniform condition ($M = 6.83°$), followed by the Scene condition ($M = 6.99°$) and the Noise condition ($M = 7.01°$). The difference between the Uniform and Scene conditions was statistically significant, $t(11) = -2.57$, $p = 0.03$, as was the difference between Uniform and Noise, $t(11) = -3.29$, $p = 0.007$. The difference between Scene and Noise was not statistically significant.

Discussion

The goal of this study was to extend results regarding the programming of saccadic responses to briefly presented targets from a static context to a dynamic context that more closely resembles the environment that saccades are programmed in when viewing naturalistic scenes. In the three conditions of the main experiment (Static, Scene, and Noise), single- or double-step targets were presented while participants were engaged in a stable fixation. In the Static task, participants responded to targets following an extended period of fixation on a central cross. During both the

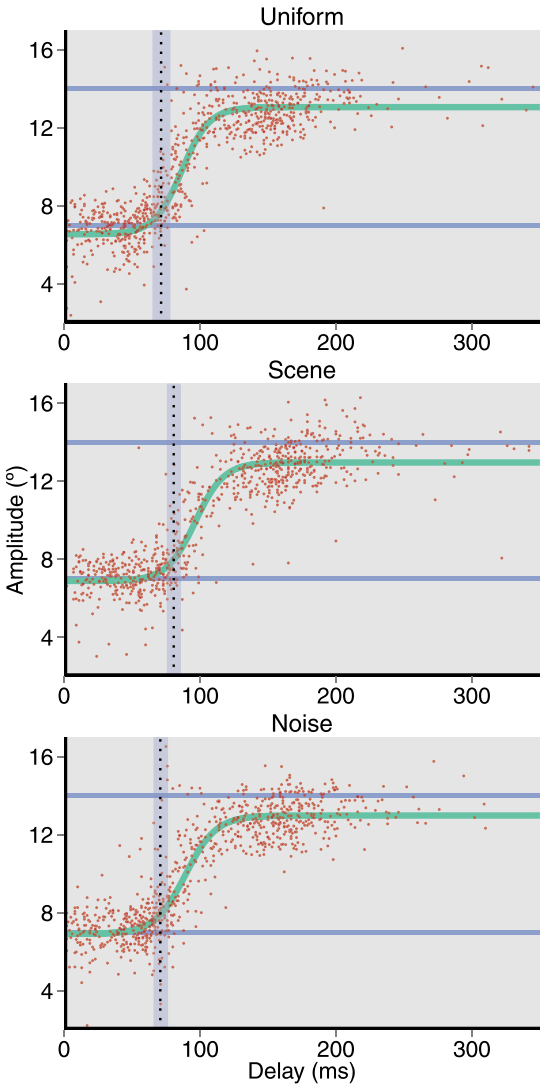

Figure 5. Amplitude transition functions for the three experimental conditions in Experiment 2. Points represent the amplitude of saccades initiated in response to the presentation of double-step targets. On the *x*-axis, the delay (D) represents the amount of time elapsed between the presentation of the second target and the onset of the saccade. The *y*-axis represents the amplitude of the resulting saccade. The green lines represent the best-fitting fixed-effects curves estimated with a nonlinear mixed-effects regression. The horizontal solid blue lines represent the physical distance of the targets in relation to fixation location (7° and 14°). The vertical dotted black lines represent the mean SDT in each condition derived from the by-participant SDT estimates. The light blue bands surrounding the means are bootstrapped confidence intervals.

Scene and Noise conditions, participants responded to targets that were presented at the onset of a fixation made while exploring the image. The double-step logic used by Becker and Jürgens (1979) was ported to the present study, which allowed comparative inferences to

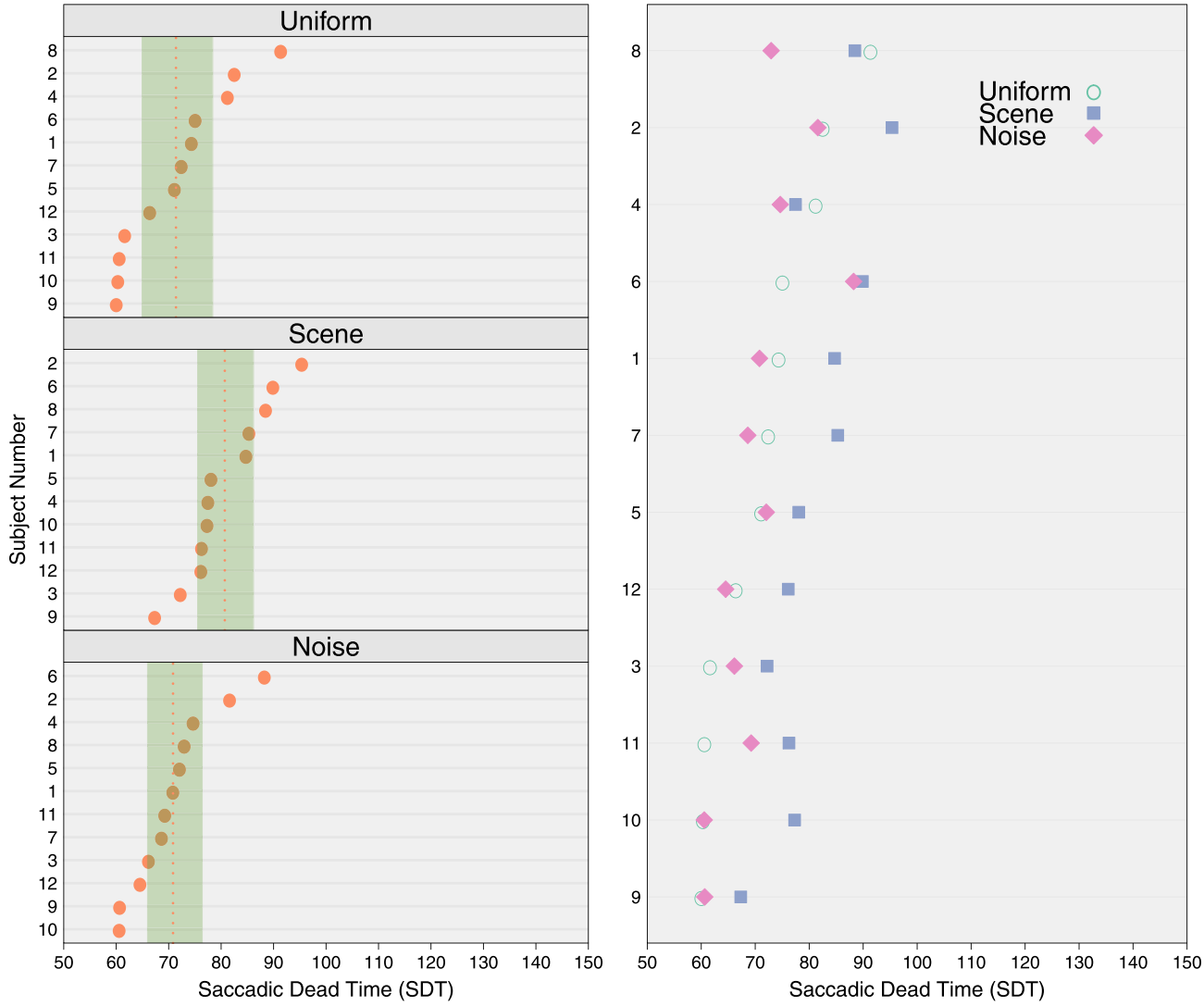

Figure 6. Saccadic dead time (SDT) estimates (ms) for the three experimental conditions in Experiment 2. The left plot shows by-participant estimates of SDT. For a given condition, from bottom to top, values are ordered from lowest to highest. The vertical dotted lines represent the mean value of SDT in each condition. The green band surrounding the mean is the bootstrapped confidence intervals of the mean estimate. The right plot shows participants ordered by the magnitude of their estimated SDT value in the uniform condition.

be made regarding the time course of saccade programming in the three conditions.

Modification of saccade programming timelines has been an influential assumption used by models that attempt to describe eye-movement control in both scene viewing and in reading (Reichle et al., 1998; Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Previously, these assumptions have primarily been warranted by double-step experiments using tasks similar to the Static condition implemented in our Experiment 1. In the present study, we observed the characteristic logistic shape of the ATF in both static and dynamic viewing conditions, which provides an important confirmation that reprogramming mechanisms that operate in static contexts operate in an analogous manner within a dynamic scene-viewing context.

ATFs permit inferences to be made regarding the time course of saccade preparation. SDT represents the point in time prior to the onset of a saccade at which that saccade may no longer be modified by updated visual information. In Experiment 1, we found that when double-step targets are presented in a static movement context in which the targets are presented on a black background, SDT is lower than it is when targets are presented in a dynamic double-step context overlaid on structured backgrounds. This difference in SDT is also complemented by a strong trend toward observing longer SDT in the Scene as compared with the Noise condition, suggesting a possible influence of

Participant	Uniform	Scene	Noise
1	74	85	71
2	82	95	82
3	62	72	66
4	81	77	75
5	71	78	72
6	75	90	88
7	72	85	69
8	91	88	73
9	60	67	61
10	60	77	61
11	60	76	69
12	66	76	65
Mean	71	81	71

Table 4. Estimated saccadic dead time (ms) in Experiment 2.

scene content per se. In Experiment 2, we isolated the influence of background on SDT by comparing three static movement conditions, which differed in the structure of the background stimulus. We found that SDT was no different in the Uniform and Noise conditions, but there was a tendency to observe a larger SDT in the Scene condition. Furthermore, by removing movement from the task, SDT estimates in the Scene and Noise conditions were greatly reduced in comparison to Experiment 1.

In Experiment 1, a notable difference between the Static and Scene task is that within the Scene task, but not the Static task, the targets are presented during a dynamic movement context. A further difference between these tasks is the high-level cognitive processes that are assumed to be activated during scene exploration. In the Scene task, participants are under instructions to view the stimulus to prepare for a later recall phase of the experiment. As a result, in the Scene condition, participants are engaged in a more complex task that is more engaging to higher-level cognition. Therefore, it might be predicted that higher-order operations specific to the processing of scene elements are implicated in the elevated SDT that was observed in the Scene condition relative to the Static condition. However, a direct comparison between these two conditions does not distinguish between (a) influences that are due to the dynamic movement and (b) differences due to additional high-level scene structure. The Noise condition in Experiment 1 was explicitly included to discriminate between these two alternatives. If scene structure does play a role in determining SDT, then it would be expected that SDT in the Scene condition is also elevated relative to the Noise condition. The pattern of results observed in Experiment 1 suggests that there may be a unique influence of scene content, as a trend toward longer SDT was observed for scene backgrounds. SDT was estimated to be 13 ms longer in the Scene compared with the Noise condition. Although this difference did not reach statistical significance, the pattern of individual results showed a clear trend toward increased SDT in the Scene as compared with the Noise condition. Results from Experiment 2 also support the view that scene content may play a role in determining SDT. As in Experiment 1, a trend for longer SDT was observed in the Scene condition relative to the non-Scene conditions. This influence on SDT occurred despite participants receiving instructions that the scene content was not relevant in any way to the successful completion of their task.

The static versus dynamic movement context also appears to play a relatively strong role in determining SDT. In Experiment 1, we observed a strong effect of movement on SDT. SDT in the two movement conditions (Scene and Noise) was elevated from the Static condition by 22 and 35 ms, respectively. As was previously argued, the SDT increase in the Noise condition relative to the Static condition should not be considered to arise from the additional structure present in the $1/f$ amplitude spectra. Experiment 2 provides support for this difference being one that is linked to movement. When movement was controlled for, no difference between the Static (Uniform) and Noise conditions was observed, and the SDT magnitudes in the Scene and Noise conditions were reduced. This was in contrast to Experiment 1 in which these backgrounds led to observed differences in SDT. This suggests that differences in SDT can arise when the eyes are in motion and actively engaged in a task, as compared to when they are at rest, waiting for stimulus presentation.

We suggest that the difference in SDT between the static and dynamic tasks arises partially due to a form of saccadic prepreparation that occurs within the Static task but is less likely to be active within the dynamic tasks. We speculate that by presenting the first target, initial stages of saccade preparation may be initiated both to the location of the visible target and to the location at which the second target is predicted to appear. In the Static task, the location of the second target is highly predictable as participants are always fixating the same location (central cross) when it appears. In contrast, estimates of the location of the target position in the dynamic task are likely to be far more variable in that there is limited evidence on which to base such predictions. As a result of such predictability, we suggest that the saccade motor system may have primed the saccade to the second target prior to the target onset. An alternative possibility is that SDT is elevated in the dynamic tasks due to the occurrence of multiple saccade plan modifications. In this account, when the first target is presented, a modification to the current saccade plan is initiated. Then, upon presentation of the second target, this updated saccade plan

must receive additional modification to reach the final target position. It may be the case that such cascaded modifications require the presence of increased SDT. As current models of saccade programming do not address such a scenario, it is difficult to make specific predictions. However, such a simulation would be an interesting and worthwhile exercise.

Stochastic process models of double-step performance in simple tasks provide an elegant framework with which to interpret how these hypothesized differences between tasks could result in SDT variability. Camalier et al. (2007) used a race model architecture to model double-step performance. In this model, saccades are generated as a consequence of a race between a saccade program targeting the first location and an additional stochastic cancellation process that is initiated upon presentation of the second target. The quantity known as the target step reaction time (TSRT) expresses the amount of time that is required by the cancellation process to successfully inhibit saccades to the first target. As lower values of TSRT correspond to lower SDT, the race model predicts that a process that results in a speed up of the cancellation process will also result in a reduction in SDT. The authors applied the race model logic to a double-step task in which the second target step competed with distractor stimuli for saccade target selection. They demonstrated that TSRT was longer when the target and distractors were more similar in color, illustrating that SDT is likely to be longer in conditions where acquisition of the second target is made more difficult. Therefore, it may be the case that lower SDT in the Static task may result from a decrease in the amount of time required for the cancellation process to complete because of the predictable characteristics of the second target location.

Ludwig et al. (2007) measured SDT in two double-step tasks. In the first task, a gap paradigm (Saslow, 1967) was used to observe SDT under conditions in which the latency of responses to targets is systematically varied. Saccade latency was manipulated by completely removing the central fixation cross prior to the onset of the saccadic response. Despite the fact that, on gap trials, saccade latency was considerably reduced (by 28%), no difference in SDT was observed. In a second static double-step experiment, the authors manipulated the angle of separation between two double-step targets such that the angle between the two targets on a given trial varied between 30° and 90°. In this task, they found a positive relationship between the degree of separation and the estimated value of SDT. They reported a mean SDT of 65 ms at the smallest separation of 30° and an SDT of 113 ms at the largest separation of 120°. The authors introduced a population coding account to explain the pattern of results. In this account, when the initial and final target are presented in close spatial proximity, there is shared activation in direction coded movement neurons for the initial and final targets. Therefore, presentation of the initial target provides a form of prepreparation for the final target movement. Because of such prepreparation, neurons coding for movement toward the second target have been partially activated and therefore require less time to reach the threshold to initiate a saccadic response. This population coding argument has been shown to adequately account for the effect of target angle separation on SDT (Ludwig et al., 2007). However, the population coding account as formulated in Ludwig et al. (2007) cannot directly account for the results reported here. In all double-step tasks, the angle between the target stimuli was maintained to be a constant 0°. Therefore, according to this framework, activation of neurons coding movement to the second target location by presentation of the first target should be equivalent in both static and dynamic tasks.

The estimated values for SDT observed in our study (Static: 74 ms, Scene: 109 ms, Noise: 96 ms) also provide a measure of empirical grounding to the values for the duration of the nonlabile stage of saccade programming that have previously been suggested (see Table 5). It is important to note that the concept of SDT does not map directly onto that of a nonlabile stage. First, the magnitude of the SDT estimates may vary depending on the method chosen to estimate the point of no return. Although a consistent method may be used to compare SDT within a single study, the magnitude of the SDT estimates may vary depending on the amplitude cutoff chosen to calculate SDT. Furthermore, any SDT estimate that is derived from double-step response data includes delays in transmission of the visual information to regions of the brain responsible for saccadic decisions. Because the nonlabile stage of saccade programming is conceptualized to operate independently of such transmission delays, deriving a nonlabile estimate from SDT requires a subtraction of this afferent delay. Neurophysiological evidence suggests that afferent delays are on the order of 50 ms (for review, see Reichle & Reingold, 2013).

Given the current method of calculating SDT, taking such delays into account would suggest that in scene viewing the duration of the nonlabile stage is approximately 60 ms. Because of the task-dependent nature of SDT, we believe that it is difficult to generalize this result to other domains. We suggest that an intriguing possibility for extension of this work would be to compare directly SDT estimates across both reading and scene-viewing tasks, thereby permitting more concrete inferences regarding the task-dependent nature of SDT. Further comment is warranted regarding the efferent delays that are known to exist in the transmission of oculomotor decisions from the brain to the eye. Such delays are thought to be on the order of

Model		Primary task	Duration (ms)	Other task	Duration (ms)
E-Z Reader					
1-5	Reichle et al. (1998)	Reading	50	—	—
9	Pollatsek et al. (2006)	Reading	25	—	—
10	Reichle et al. (2012)	Reading	25	Search	25
10	Reichle et al. (2012)	Reading	25	Z-String	25
SWIFT					
I	Engbert, Longtin, and Kliegl (2002)	Reading	41.6	—	—
II	Engbert et al. (2005)	Reading	[6.1, 58.7]	—	—
II	Nuthmann and Engbert (2009)	Reading	[6.1, 58.7]	Z-String	[20.5, 59.6]
III	Schad and Engbert (2012)	Reading	[50, 51.3]	Shuffled	[50, 51.3]
CRISP					
	Nuthmann et al. (2010)	Scene viewing	40	—	—
	Nuthmann and Henderson (2012)	Scene viewing	40	Reading	14
ICAT					
	Trukenbrod and Engbert (2014)	Search	40	Reading	40

Table 5. Comparison of model nonlabile estimates. *Notes:* As of SWIFT-II, the model predicts a continuum of nonlabile durations. A range of values are provided that reflect short (1°) and long (10°) reading saccades.

20 ms (Becker, 1991). In the saccade programming architecture of the models that we have discussed, such delays must logically occur toward the latter stages of the nonlabile stage. The duration of the efferent delay places a lower bound on the duration of the nonlabile stage as no stimulus information may contribute to the programming of a saccadic response after the movement signal has been sent to the motor effectors.

The present results are particularly informative for models of scene perception (Nuthmann et al., 2010), reading (Engbert et al., 2005; Reichle et al., 1998), and visual search (Trukenbrod & Engbert, 2014) that use multistage saccade programming architectures. Table 5 shows the parameters used to specify the mean duration of the nonlabile stage of saccade programming in these models. Surprisingly, there is little consistency with regard to the duration of the nonlabile stage within a given task. In some cases, the models predict durations that are less than the minimum interval required by the efferent delay. This represents a potential difficulty in that such models aim to synthesize empirical knowledge regarding the timeline of oculomotor control into predictions about behaviors in high-level tasks. One of the contributions of the present study is to provide a principled experimental grounding for assumptions that are critical to such behavioral predictions. However, it is important to note that differences between the double-step tasks under investigation here and other typical scene-viewing contexts do exist. For example, in the dynamic viewing tasks, participants are aware that on each trial, a sudden target onset will occur after the scene has been explored for some number of fixations. Consequently, it is possible that toward the later stages of a trial, participants may begin to modify their viewing strategies. It is therefore worthwhile considering that these tasks are designed to study saccade programming during scene perception tasks that include a target acquisition component.

There is also some discord between models on the question of task-dependent differences in the duration of the nonlabile stage. In both the E-Z Reader model (Reichle et al., 2012) and the CRISP model (Nuthmann et al., 2010), there is no mechanism provided by which the mean nonlabile duration may vary within a specific task.[2] In Reichle et al. (2012), an estimated mean duration of 25 ms remained constant both within and across reading, z-string reading, and search tasks. In simulations with the CRISP model (Nuthmann & Henderson, 2012), a comparison was made between model predictions of fixation durations in reading and in scene viewing. In these simulations, the duration of the nonlabile stage was permitted to vary across tasks (reading: 14 ms; scene viewing: 40 ms). Similarly to the CRISP model, the SWIFT model also predicts task-dependent differences in mean nonlabile durations (Nuthmann & Engbert, 2009). However, as of SWIFT-II, the model additionally assumes that nonlabile durations vary systematically within a task by scaling the nonlabile duration to the length of the resulting saccade (Engbert et al., 2005). SWIFT-II was evaluated with data from the Potsdam Sentence Corpus, in which one letter subtends 0.38° and/or 0.45° of visual angle (Nuthmann & Kliegl, 2009). For short saccades ($\approx$ 1°) SWIFT-II predicts average nonlabile durations as long as 58.7 ms. At the other end of the continuum, long saccades ($\approx$ 10°) are predicted to have an average nonlabile duration of 6.1 ms. Therefore, between-task differences in average saccade amplitude may translate into differences in nonlabile durations. However, it should also be noted that estimated parameters in a later version of the SWIFT model result in a model with no between- or within-task variability in nonlabile dura-

tions (Schad & Engbert, 2012). A recent model, ICAT (Trukenbrod & Engbert, 2014), has modeled fixation durations in both visual search and reading tasks. In ICAT, a fixed nonlabile duration of 40 ms was assumed for both tasks. Future modeling efforts would greatly benefit from addressing the question of what assumptions are being made with regard to the nonlabile values that are selected or estimated from the data.

Conclusions

Computational models have played an important role in our understanding of oculomotor control behavior in high-level tasks such as scene viewing and reading. However, these models have often relied on simple stimuli used in basic psychophysical paradigms to inform the development of their architectures. Here, using a gaze-contingent double-step paradigm, we have generalized these findings to a more ecologically valid context and revealed important task differences in saccade programming. The present results are particularly informative for the understanding of saccade programming during scene viewing. However, we suggest that future work should directly investigate the processes of saccade cancellation in reading and other task contexts. These studies would provide further generalization for a role of saccade cancellation in eye-movement control, as well as provide empirical validation for task-specific modeling efforts.

Keywords: saccade programming, scene viewing, double-step, eye-movement control, reading

Acknowledgments

Commercial relationships: none.
Corresponding author: R. Calen Walshe.
Email: calen.walshe@gmail.com.
Address: Psychology Department, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

Footnotes

[1] We thank Casimir Ludwig for suggesting this method. We also note that calculating SDT in such a manner resulted in estimates that were increased in comparison to an alternative method (see Ludwig et al., 2007).

[2] Although there is no systematic variability built into the nonlabile duration, stochastic variability does enter these models via the duration being sampled from a gamma distribution.

References

Aslin, R. N., & Shea, S. L. (1987). The amplitude and angle of saccades to double-step target displacements. *Vision Research, 27*, 1925–1942.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye-movements: Vol. 8 in vision and visual dysfunction* (pp. 95–137). Boston, MA: CRC Press.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research, 19*, 967–983.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research, 47*, 2187–2211.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*, 613–617.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Einhäuser, W., Rutishauser, U., Frady, E. P., Nadler, S., König, P., & Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. *Journal of Vision, 6*(11):1, 1148–1158, http://www.journalofvision.org/content/6/11/1, doi:10.1167/6.11.1. [PubMed] [Article]

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2):2, 1–19, http://www.journalofvision.org/content/8/2/2, doi:10.1167/8.2.2. [PubMed] [Article]

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.

Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 71–78). Amsterdam: Elsevier.

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research, 48*, 1777–1790.

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 318–322.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*, 566–573.

Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision, 12*(2):4, 1–13, http://www.journalofvision.org/content/12/2/4, doi:10.1167/12.2.4. [PubMed] [Article]

Kaspar, K., & König, P. (2011). Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. *PLoS One, 6*, e21719.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 14.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559–3565.

Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology, 97*, 795–805.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8):17, 1–15, http://www.journalofvision.org/content/11/8/17, doi:10.1167/11.8.17. [PubMed] [Article]

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences, 1339*, 82–96.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research, 49*, 322–336.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19, http://www.journalofvision.org/content/10/8/20, doi:10.1167/10.8.20. [PubMed] [Article]

Nuthmann, A., & Henderson, J. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition, 20*, 457–494.

Nuthmann, A., & Kliegl, R. (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision, 9*(5):31, 1–28, http://www.journalofvision.org/content/9/5/31, doi:10.1167/9.5.31. [PubMed] [Article]

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*, 382–405.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2014). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1–117. Vienna, Austria: R Foundation for Statistical Computing.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*, 1–56.

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ray, S., Schall, J. D., & Murthy, A. (2004). Programming of double-step saccade sequences: Modulation by cognitive control. *Vision Research, 44*, 2707–2718.

Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125–157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review, 119*, 155–185.

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience, 7*, 1–6.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65*, 177–206.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America, 57*, 1024–1029.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*, 391–421.

Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics, 53*, 305–314.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8):16, 1–24, http://www.journalofvision.org/content/13/8/16, doi:10.1167/13.8.16. [PubMed] [Article]

SR Research Ltd. (2006). *EyeLink programmer's guide*. (3.0 ed.). Retrieved from download.sr-support.com/dispdoc.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2):5, 1–18.

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review, 21*, 904–934.

Walshe, R. C., & Nuthmann, A. (2013). Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1569–1574). Berlin. Cognitive Science Society.

Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research, 100*, 38–46.

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology, 52*, 932–941.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology, 9*, e1002871.

Wu, E. X. W., Gilani, S. O., van Boxtel, J. J. A., Amihai, I., Chua, F. K., & Yen, S.-C. (2013). Parallel programming of saccades during natural scene viewing: Evidence from eye movement positions. *Journal of Vision, 13*(12):17, 1–14, http://www.journalofvision.org/content/13/12/17, doi:10.1167/13.12.17. [PubMed] [Article]

C

APPENDIX C

Vision Research 100 (2014) 38–46

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Asymmetrical control of fixation durations in scene viewing

R. Calen Walshe *, Antje Nuthmann

Psychology Department, University of Edinburgh, Edinburgh, UK

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form 3 March 2014
Available online 12 April 2014

Keywords:
Fixation durations
Eye movements
Scene viewing
Direct control

ABSTRACT

In two experiments we investigated the control of fixation durations in naturalistic scene viewing. Empirical evidence from the scene onset delay paradigm and numerical simulations of such data with the CRISP model [Psychological Review 117 (2010) 382–405] have suggested that processing related difficulties may lead to prolonged fixation durations. Here, we ask whether processing related facilitation may lead to comparable decreases to fixation durations. Research in visual search and reading have reported only uni-directional shifts. To address the question of unidirectional (slow down) as opposed to bidirectional (slow down and speed up) adjustment of fixation durations in the context of scene viewing, we used a saccade-contingent display change method to either reduce or increase the luminance of the scene during prespecified critical fixations. Degrading the stimulus by shifting luminance down resulted in an immediate increase to fixation durations. However, clarifying the stimulus by shifting luminance upwards did not result in a comparable decrease to fixation durations. These results suggest that the control of fixation durations in scene viewing is asymmetric, as has been reported for visual search and reading.

1. Introduction

The study of eye guidance during naturalistic scene viewing aims to understand the processes that underlie the acquisition of vital visual information from the environment that is relevant to current tasks and goals. Described in a very general manner, investigation into the control of eye movements in scene viewing has proceeded along two primary pathways. The first seeks to address questions relating to where eye movements are directed towards, while the second addresses questions regarding when the eyes move away from currently fixated content. The first question, relating to the spatial aspects of eye movements, has received considerable attention while there is relatively less research investigating the related temporal component (Murray, Fischer, & Tatler, 2013). Mean fixation durations in scene viewing are about 300 ms (Rayner, 2009) but there is considerable variability around this mean both within and across individuals. Current understanding of eye-movement programming suggests that some of the variability in the duration of individual fixations may result from factors directly related to oculomotor programming (Becker & Jürgens, 1979; Nuthmann et al., 2010; Walshe & Nuthmann, 2013), as well as global scene properties (e.g., Loftus, 1985; Henderson, Nuthmann, & Luke, 2013; Nuthmann et al., 2010), and decisional processes relating to future target selection (Glaholt & Reingold, 2012).

The structure of the mechanisms that govern fixation times has been investigated in a wide variety of tasks (Rayner, 2009). Research that addresses these questions often aims to reveal the manner in which the eye-movement control system adaptively monitors and responds to environmental demands. A debate of critical importance for the understanding of the temporal characteristics of fixation times is the degree to which stimulus content that is currently under inspection influences the decision of when to terminate the current fixation (Reingold et al., 2012). The hypothesis that fixations are capable of being adjusted on a moment-to-moment basis is referred to as the direct control hypothesis (reading: Rayner & Pollatsek, 1981, scene perception: Henderson & Pierce, 2008; Nuthmann et al., 2010).

This hypothesis is characterised by the assertion that when a fixation duration is under the direct control of stimulus content, there is an immediate adjustment to match the processing demands of the stimulus. In contrast, fixations may be indirectly controlled, and this occurs in the case where fixation times are governed by influences that extend beyond the locally fixated content. For instance, from studies of visual search it is known that fixation durations increase as the complexity of the search array increases (Vlaskamp & Hooge, 2006), when target-distractor similarity is increased (Hooge & Erkelens, 1998; Vlaskamp et al., 2005), and in order to match the difficulty of previously fixated items (Hooge & Erkelens, 1998). These results imply that the

* Corresponding author. Address: Psychology Department, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
E-mail addresses: r.c.walshe@sms.ed.ac.uk (R. Calen Walshe), antje.nuthmann@ed.ac.uk (A. Nuthmann).

http://dx.doi.org/10.1016/j.visres.2014.03.012
0042-6989/

eye-movement control system is sensitive, at least in some part, to the global characteristics of the task.

A variety of direct-control mechanisms have been proposed to account for the moment-to-moment adaptation of fixations to current stimulus processing. Concepts related to the structure of direct control mechanisms have seen the most development in theories of fixation times in reading. In reading, a debate exists regarding how the lexical properties of the currently fixated word impacts the time course of that fixation. Mechanisms used to account for such lexical effects may be contrasted as those that implement what is known as a *cognitive trigger*, and those that implement *interference* mechanisms (see Reingold et al., 2012). Cognitive trigger theories postulate that the decision to terminate a fixation is made once the stimulus under inspection has been processed to a sufficient degree, and when this occurs a saccade programme is then triggered. One implementation of such a mechanism is incorporated in the E-Z Reader model, in which an eye-movement programme is triggered once a superficial stage of lexical processing has been accomplished (Reichle et al., 1998; Reichle, Pollatsek, & Rayner, 2012). In contrast to the triggering mechanisms just described are those that suggest that the variability in the termination of a fixation is a result of difficulties in lexical processing that interfere with the saccade initiation processes. A model that instantiates a variety of direct control along these lines is the SWIFT model (Engbert et al., 2005). In the SWIFT model, the decision to initiate a saccade programme is achieved by an autonomous random timer, and the duration of this timing process may be modulated by the difficulties encountered during lexical processing. Therefore, moment-to-moment difficulties in lexical processing results in increased random timing intervals, and consequently, longer fixation durations.

Although less is known about the mechanisms that govern eye-movement control in scene perception, a model that incorporates an interference mechanism to explain fixation times in this domain is known as the CRISP model (Nuthmann et al., 2010). In this model, an autonomous random walk timer accumulates towards a fixed threshold value and when this threshold is reached, a saccade program is initiated. In the case in which processing difficulties are encountered during scene viewing, the rate at which the timer accumulates to the threshold is reduced. A consequence of such a reduction in the rate of the timer is that the initiation of saccades may be delayed, and therefore longer fixation durations will be observed. An assumption that was made in the original formulation of the CRISP model was that modulations to the timer result exclusively from unidirectional modulations (timer slowdown) (Nuthmann et al., 2010).

An experimental paradigm that has provided some evidence for the direct control of fixations in scene viewing is known as the scene onset delay (SOD) paradigm (Henderson & Pierce, 2008; Henderson & Smith, 2009; Luke, Nuthmann, & Henderson, 2013; Nuthmann et al., 2010; Nuthmann & Henderson, 2012). In the SOD paradigm, a scene is masked during a saccade preceding a critical fixation and then restored to full view at varying delays within the critical fixation. Consistently across studies, a population of fixation durations that increased in correspondence with the length of the delay was observed. It was argued that these fixations were increased due to the immediate effects attributable to the missing stimulus. Pannasch, Schulz, and Velichkovsky (2011) used a scene based free viewing task in which an irrelevant distractor was introduced either early or late within a critical fixation. Similar to the SOD paradigm, the distractors were presented for variable durations. The results demonstrated that the visual change introduced by the distractor had an immediate prolongation effect on fixation durations, regardless of whether the distractor occurred early or late in fixation, which provided additional support for the direct-control hypothesis.

Going beyond the extreme manipulations of the SOD paradigm, subsequent research has utilised a fixation-contingent scene quality paradigm (Henderson, Nuthmann, & Luke, 2013; Glaholt, Rayner, & Reingold, 2013). During selected critical fixations, the entire scene was reduced in quality via a decrease in luminance (Henderson, Nuthmann, & Luke, 2013), or by filtering high or low spatial frequencies (Glaholt, Rayner, & Reingold, 2013). Such manipulations are assumed to have deleterious effects on scene processing by influencing the rate at which information is extracted from scenes (Loftus, 1985) as well as impacting the fluent encoding of scene stimuli into working memory (Glaholt, Rayner, & Reingold, 2013). In a study by Henderson, Nuthmann, and Luke (2013), the luminance of the (colour) scene was reduced during the saccade prior to a prespecified critical fixation. During the saccade that terminated the critical fixation, the scene returned to its normal luminance. The durations of the critical fixations were immediately affected by the reduction in scene luminance, with increasing durations for decreasing luminance. Glaholt, Rayner, and Reingold (2013), on the other hand, demonstrated that fixation durations were affected on a fixation-by-fixation basis depending on the spatial frequency content of the scene stimulus. In their main experiment, during the critical fixation the (greyscale) scene was changed to a high-pass or low-pass spatial frequency filtered version. Under both conditions, fixation durations increased, and low-pass filtering produced a greater effect than high-pass filtering. In a further experiment, the authors additionally modified the orientation of the images, and using a distributional analysis of fixation durations they were able to differentiate between directly controlled extensions to fixations attributable due to higher-level cognitive influences, and transsaccadic changes resulting in a surprise effect. These results taken together, are highly suggestive that in scene viewing, as in reading, the control of fixation durations is subject to ongoing visual-cognitive processing, such that increases to processing difficulty result in extended fixation durations.

However, further questions regarding the properties of this direct-control process remain. For instance, in the studies that were previously reviewed, the observed effects on fixation durations were primarily ones in which an increase in processing difficulty resulted in an extension to fixation durations. Therefore, these studies demonstrate that there is a tendency for fixations to be immediately adjusted to match the difficulty of the stimulus. However, it is less clear whether the converse is true. That is, will a decrease to fixation durations be observed in the case in which the processing of a stimulus becomes easier and more fluent?

In reading, Kennison and Clifton (1995) investigated the impact of word frequency on two adjacent words embedded in single sentences. High and low word frequency adjectives were followed by high and low word frequency nouns. Parafoveal preview of the noun was prevented by using the invisible boundary technique. When readers first fixated a high-frequency adjective, fixation durations on the subsequent noun showed a word frequency effect, such that longer fixation durations were observed for low-frequency than for high-frequency nouns. In contrast, no such word frequency effect was observed when readers first fixated a low-frequency adjective. Thus, increasing processing demands (high → low) resulted in an immediate prolongation of fixation durations, whereas decreasing processing demands (low → high) showed no immediate facilitatory effect.

Such an asymmetry in the temporal control of fixation durations has also been observed in visual search. Hooge, Vlaskamp, and Over (2007) used a search task in which participants were required to find a closed ring amongst distractor Cs. The distractors in their task varied in the size of the gap, such that small gap Cs were more difficult to distinguish from the target stimulus than were large gap Cs. They found that fixations on small gap Cs that

were preceded by a fixation on a large gap C showed increased durations. However, a fixation on a large gap C following a fixation on a small gap did not show a corresponding decrease to fixation duration. These results taken together suggest that the control of fixation durations in both reading and visual search tasks involves an asymmetrical pattern of control. While these results provide some guidance on the question of whether asymmetrical control principles generalise to scene viewing tasks, there currently exists no experimental evidence to confirm whether this is the case.

The purpose of the current study was to directly test the hypothesis that the control of fixation durations in scene viewing is asymmetric. To manipulate processing difficulty of the currently fixated stimulus, the present study employed a luminance manipulation such that increased difficulty was obtained by shifting luminance downwards, and decreased difficulty was obtained by shifting luminance upwards. The assumption that modulation of scene luminance levels may be used to control the difficulty of scene processing is derived from several sources. Past research has shown that luminance has strong effects on scene processing, with lowered recognition and recall rates of scenes when they are viewed at a lower level of luminance (Loftus, 1985; van der Linde et al., 2009). These effects are paralleled by an increase in fixation durations to compensate for the increase in processing difficulty encountered due to the luminance reduction (Loftus, 1985). More recently, a control experiment conducted by Henderson, Nuthmann, and Luke (2013) used a free viewing task in which scenes were viewed at 100%, 80%, or 60% original scene luminance throughout the course of the entire trial. They demonstrated that scene luminance had a clear influence on fixation durations such that longer mean fixation durations were observed when scenes were viewed at lower luminance levels. Therefore, these results taken together support the assumption that scene luminance is parametrically related to scene processing difficulty.

In order to test the hypothesis that the direct control mechanism operates in an asymmetric manner, a fixation-contingent scene quality paradigm was used (Henderson, Nuthmann, & Luke, 2013). With this method, the luminance shifts took place during saccades when visual transients were suppressed (Ross et al., 2001; McConkie & Loschky, 2002). While it may be predicted from the gaze-contingent manipulations of Henderson, Nuthmann, and Luke (2013), that longer fixation durations will be observed following a gaze contingent decrease in luminance, it is currently unclear how an increase in luminance will be interpreted by the eye-movement control system during naturalistic scene viewing. The prediction of an asymmetrical direct-control mechanism is that decreased luminance will result in longer fixations, while increased luminance will have no effect. In contrast, symmetrical direct control would predict that shifting luminance down will result in longer fixation durations, and clarifying the stimulus by shifting luminance up will result in shorter fixation durations.

2. Experiments

2.1. General methods

2.1.1. Stimuli

In each of two experiments, participants viewed a total of 100 pictures of real-world scenes, in addition to 4 practice scenes. Each scene had a resolution of 800x600 pixels and was presented in full colour. Scenes were collected from online databases such as google images. They were selected to include a variety of categories such as indoor and outdoor as well as urban and nature scenes. Each scene was viewed by the participant only once over the duration of the experiment, and the experimental scenes were presented to the participants in a randomised order. Initially, scenes were presented at a baseline luminance of 80% in Experiment 1, and 60% in Experiment 2. In order to observe the effect of relative luminance shifts on fixation durations, a luminance transformation was applied. Luminance shifted stimuli were created by converting the original scene into a $L*a*b$ colour space (Oliva & Schyns, 2000), and modifying the luminance channel L by the appropriate value. This procedure allows the separation of a luminance channel from the two colour channels, and permits the transformation of scene luminance independently of scene colour. Baseline and low luminance conditions for Experiment 1 were constructed by a $L*.8$ and $L*.6$ transformation, respectively. For Experiment 2, a similar procedure was adopted, but the luminance transformation applied was $L*.6$ and $L*.2$. In both Experiment 1 and Experiment 2, the stimulus used in the high (100%) luminance condition was the untransformed scene.

2.1.2. Procedure

Participants were instructed that they would take part in an experiment in which they would see many pictures of naturalistic content and that their task was to encode the scenes for later recall. They were instructed that the recall phase would only begin once all the scenes had been viewed, but were not told how many scenes would be presented. These instructions were provided only to motivate scene encoding behaviour, and therefore the recall phase was not applied. Following the instructions, a nine-point eye-tracker calibration procedure was initiated. A trial began when the participant fixated on a cross presented at the centre of the screen. Following this fixation, the red cross and grey background were replaced with the scene presented at baseline luminance. Participants then engaged in the encoding task until a critical fixation was identified when a participant had made at least 10 saccades since the beginning of the trial. If a critical fixation had been identified, the luminance shift was made during the saccade immediately preceding the critical fixation. The luminance-shifted scene was presented for the entire duration of the critical fixation, and the luminance was then shifted back to baseline during the saccade immediately following the critical fixation. In total, four luminance manipulations were made on each trial; two manipulations resulted in an upward luminance shift, and two manipulations were made in the downward direction. After the first luminance manipulation had been completed, subsequent shifts occurred on every 10th saccade since the most recent luminance shift. The order of the luminance shift direction (increase vs. decrease), was randomised within a trial. Once the fourth luminance shift had been made, and the participant terminated the resulting critical fixation, one second elapsed until the trial was terminated. The scene was then replaced with a grey background and red fixation cross. Once the participant fixated on the cross, the next trial was initiated. In the situation that the trial lasted longer than 25 s, the current trial was abandoned, and the participant was presented with a fixation cross to initiate the next trial. A schematic of the procedure for upward luminance shifts is presented in Fig. 1. The mean trial length in Experiment 1 was 18.1 s and 19.2 s in Experiment 2. The mean number of saccades per trial was 50.3 in Experiment 1, and 48.1 in Experiment 2.

2.1.3. Apparatus

Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz. The monitor screen was at a distance of 90 cm from the participant. During stimulus presentation, participants' eye movements were recorded using an SR Research EyeLink 1000 Desktop mount system. It was equipped with the 2000 Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz for each eye. Viewing was binocular, and both eyes were tracked. A chin rest was used in order to achieve stability of a participant's head position relative to the screen. The experiment was

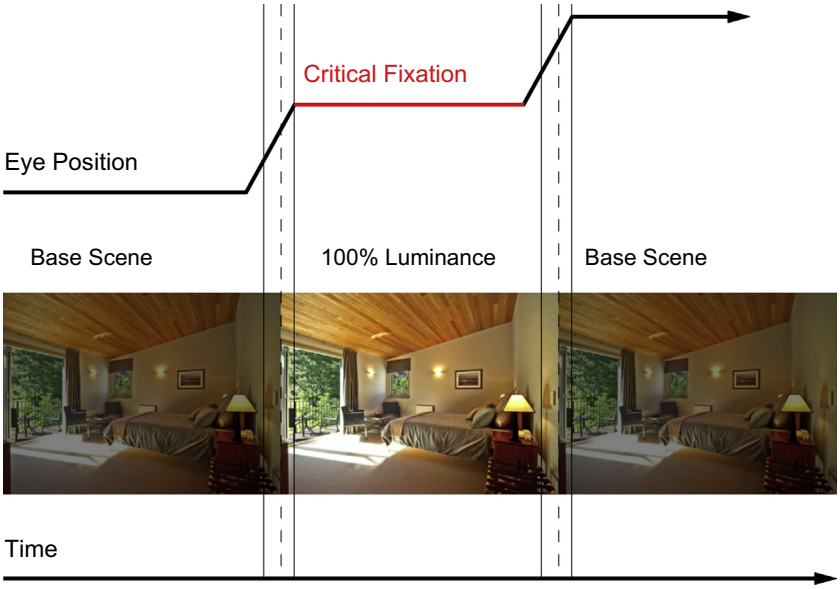

Fig. 1. A schematic of the paradigm used to create gaze contingent luminance shifts. Base scenes represent the image that is viewed during the fixation immediately preceding a critical fixation. A critical fixation is defined to occur on the 10th fixation since the previous luminance manipulation. The oblique lines represent saccadic eye movements. During a saccadic eye movement, the scene is either increased or decreased in luminance. A critical fixation is terminated upon detection of a saccadic eye-movement, and the scene is restored to base scene luminance during this saccade.

implemented in MATLAB 2009b using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), which incorporates the EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002). The software allowed precise control over the timing of display changes.

Online detection of saccades involves a speed-accuracy trade off, such that incorporating more samples reduces the noise in the signal. However, by increasing the number of samples, measurement lag is increased, which decreases the temporal precision with which saccades are detected. We implemented a 9-sample online velocity detection algorithm in MATLAB that aimed to mimic Data Viewer's offline saccade detection procedure (SR Research Ltd, 2006). Saccades were identified when gaze data from the right eye reached a two-dimensional velocity threshold of 85°/*s*. Raw data was post-processed utilising SR Research Data Viewer to parse the gaze samples into sequences of fixations and saccades.

Several data exclusion criteria were applied to remove critical fixations that had been misidentified. Prior to any data exclusion, 97.9% of the luminance manipulations were executed in Experiment 1 and 95.7% were executed in Experiment 2. This number is less than 100%, as a trial would occasionally timeout before all luminance shifts had been completed. Critical fixations on which the display change did not complete prior to fixation onset were discarded. This criteria was validated by comparing the saccades detected online with saccades identified by the post-processed Data Viewer output. Comparison with the post-processed data represents an objective measure, as this data incorporates acceleration and velocity of both prior and future eye-position samples, in detecting current saccadic activity. This resulted in retention of 85.4% of the data in Experiment 1, and 86.4% in Experiment 2. Critical fixations that co-occurred with blinks were also excluded from the analysis. Removing blinks resulted in 67.5% of the critical fixations being retained in Experiment 1, and 68.4% in Experiment 2. A final criteria was applied that excluded critical fixations that had durations of less than 50 ms or greater than 1200 ms, on the assumption that they are not determined by cognitive level processes under investigation in this study (Inhoff & Radach, 1998). As a result of the application of all criteria, 65.8% of the critical fixation were retained in Experiment 1 and 65.1% were retained in Experiment 2.

2.1.4. Analysis

Data were analysed with linear mixed-effects (LME) models, using the `lmer` programme of the `lme4` package (Bates, Maechler, & Bolker, 2012) implemented in the R statistical computing software (R Core Team, 2012). To evaluate the effect of the downward and upward luminance shifts on fixation duration, we used treatment contrasts in which the baseline condition, where no luminance change occurred, served as the reference group. Consequently, the intercept for the fixed effect "luminance shift", estimates the mean value for the no-shift condition. The two slopes estimate the difference between downward luminance shift and no shift (DOWN) and between upward luminance shift and no shift (UP). The effect of luminance is assessed in the LME model by observing regression coefficients for the luminance shift conditions that are significantly different from 0; a two-tailed criterion of $t = 1.96$ was used to assess statistical significance. The LME models included random intercepts and random slopes for participants and items (Baayen, Davidson, & Bates, 2008).

Additional ex-Gaussian distributional analyses of fixation durations were conducted by employing a generalised additive model location, scale and shape (GAMLSS) framework, using the `gamlss` package (Rigby & Stasinopoulos, 2005) implemented in R. GAMLSS is a regression framework that allows the response variability to be modelled by skewed distributions such as the ex-Gaussian distribution. Regression coefficients of the ex-Gaussian parameters contrasted the two treatment conditions (DOWN and UP) with the baseline condition. A two-tailed criterion of $t = 1.96$ was used to assess statistical significance.

2.2. Experiment 1

2.2.1. Methods

Stimuli. The stimuli used in Experiment 1 were presented at a baseline level of 80% of original scene luminance throughout the trial. Upon detection of a saccade preceding the critical fixation, the stimulus was replaced with a scene which had the luminance

raised or lowered by a margin of 20%. This meant that in the DOWN condition, participants viewed a stimulus at 60% original luminance, and in the UP condition participants viewed a stimulus at 100% original luminance. During the saccade that terminated the critical fixation, the scene returned to its base luminance.

Participants. Four males and 18 females were recruited from the University of Edinburgh student population. The mean age of the participants was 21 years. Each participant was paid £7 per hour of participation in compensation for their time.

2.2.2. Results

The goal of the analysis was to assess the impact that gaze-contingent luminance shifts have on fixation durations. Therefore, our analysis was restricted to critical fixations that began following the termination of a saccade and ended with the initiation of a subsequent saccade. In all cases, the critical fixation was defined such that a luminance manipulation had been made during the saccade immediately preceding the fixation. A baseline measure was constructed in order to detect differences between luminance shifted fixations and fixations in which no luminance shift took place. For each luminance manipulation that survived the exclusion criteria, we measured the duration of the fixation immediately preceding the critical saccade. Since the participant was unaware that a luminance manipulation was to take place during the subsequent saccade, this fixation duration represents an independent measure of fixation duration on the unmodified image. It is important to note that a baseline condition with a greater number of observations than were present in either the UP or the DOWN condition was used (cf., Glaholt, Rayner, & Reingold, 2013). A strength of the linear mixed-effects modelling approach adopted in the present study is that it is known to yield robust parameter estimates for such unbalanced designs (Baayen, Davidson, & Bates, 2008).

The mean pattern of critical fixation durations is presented in Fig. 2. To reiterate, the intercept for the fixed effect of luminance shift estimates the mean value for the no-shift condition ($b = 297.30, SE = 9.20, t = 32.30$). As expected, downward luminance shifts were associated with critical fixations that were

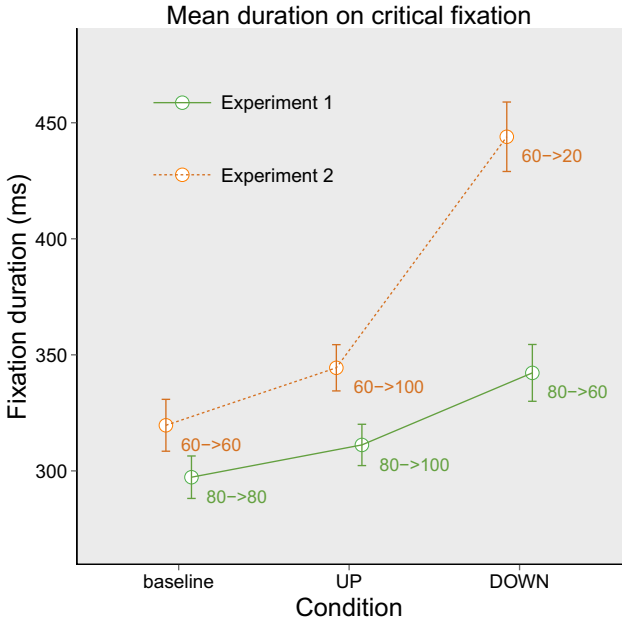

Fig. 2. Mean fixation durations on critical fixations following gaze-contingent luminance shifts. Fixation durations are plotted as a function of the direction of luminance shift. Data is plotted for Experiment 1 (solid line) and for Experiment 2 (dashed line). Error bars represent the standard error of the mean.

significantly longer than in the no-shift condition ($b = 44.92, SE = 6.07, t = 7.40$). In addition, there was also a significant increase in fixation durations for upward luminance shifts ($b = 13.28, SE = 4.21, t = 3.15$). The effect of the UP condition is contrary to predictions by both the asymmetric control hypothesis (no change) and symmetric control hypothesis (decrease). When translating the estimated effects of luminance shift into a % increase relative to baseline, it becomes apparent that the effect was much smaller in the UP condition (4.5% increase) than in the DOWN condition (15.1% increase). Comparing the between condition means is informative for the asymmetrical control hypothesis under investigation in the current study. However, changes in mean fixation duration (or the lack thereof) may reflect distinct patterns at the level of underlying distributions. More specifically, previous work on eye guidance in reading and scene perception has argued that applying an ex-Gaussian distributional analysis to fixation durations allows inferences about the time course of effects by quantifying whether effects may be attributed to a shift in central tendency or tail of the distribution (Glaholt and Reingold, 2012; Reingold et al., 2012; Luke, Nuthmann, and Henderson, 2013; Staub et al., 2010). The ex-Gaussian is a three-parameter distribution that is derived by a convolution of the Gaussian distribution with the exponential distribution. The parameters contributed by the Gaussian distribution are μ and σ, and describe the central tendency and the spread of the distribution. The τ parameter contributed by the exponential distribution provides a measure of the skewness of the distribution and is useful for describing effects that specifically impact the tail of the distribution.

Fig. 3a and c plot the empirical distribution and ex-Gaussian fitted distributions for Experiment 1. Consistent with the findings from the analysis of means, the distributions for both the DOWN and UP condition are shifted to the right relative to the baseline condition, indicating a higher probability of observing longer fixation durations in these conditions. Accordingly, there was a significant effect of DOWN on μ in the ex-Gaussian fit ($b = 50.06, SE = 2.87, t = 17.40$). There was also a significant effect of UP on μ ($b = 32.04, SE = 2.64, t = 12.11$). A statistically significant effect of DOWN on σ was observed ($b = 10.29, SE = 2.11, t = 4.87$), indicating that the spread of the distribution was larger than in the no-shift control condition. In contrast, there was no effect on σ for the UP condition ($b = 1.59, SE = 1.99, t = 0.79$). An analysis of the τ parameter of the ex-Gaussian fit revealed that the increase in fixation durations in the DOWN and UP conditions does not result from increases that are specifically attributable to the tails of the distributions. Rather, the fixation duration distribution in the UP condition was significantly less skewed than the baseline condition, evidenced by a significant negative effect on τ ($b = -19.92, SE = 3.86, t = -5.13$). DOWN had a small, marginally significant, negative effect on τ ($b = -8.08, SE = 4.14, t = -1.94$).

2.2.3. Discussion

A 20% luminance reduction of the entire scene during critical fixations was associated with an immediate lengthening of those fixations' duration. The pattern of mean fixation durations for a fixation-contingent downward shift in luminance is consistent with results by Henderson, Nuthmann, and Luke (2013). Thus, we provide a replication of their results with a different base luminance level (80% rather than 100%), different scene stimuli and participants, and statistical evaluation that controlled for variability introduced by participants and items. In addition, the current experiment included a condition in which processing was made easier by shifting luminance upwards (from 80% to 100%). There was no facilitatory effect of shortened fixation durations observed in this condition, which is consistent with research in visual search (Hooge, Vlaskamp, & Over, 2007) and in reading (Kennison & Clifton, 1995). On the contrary, in the UP condition we observed

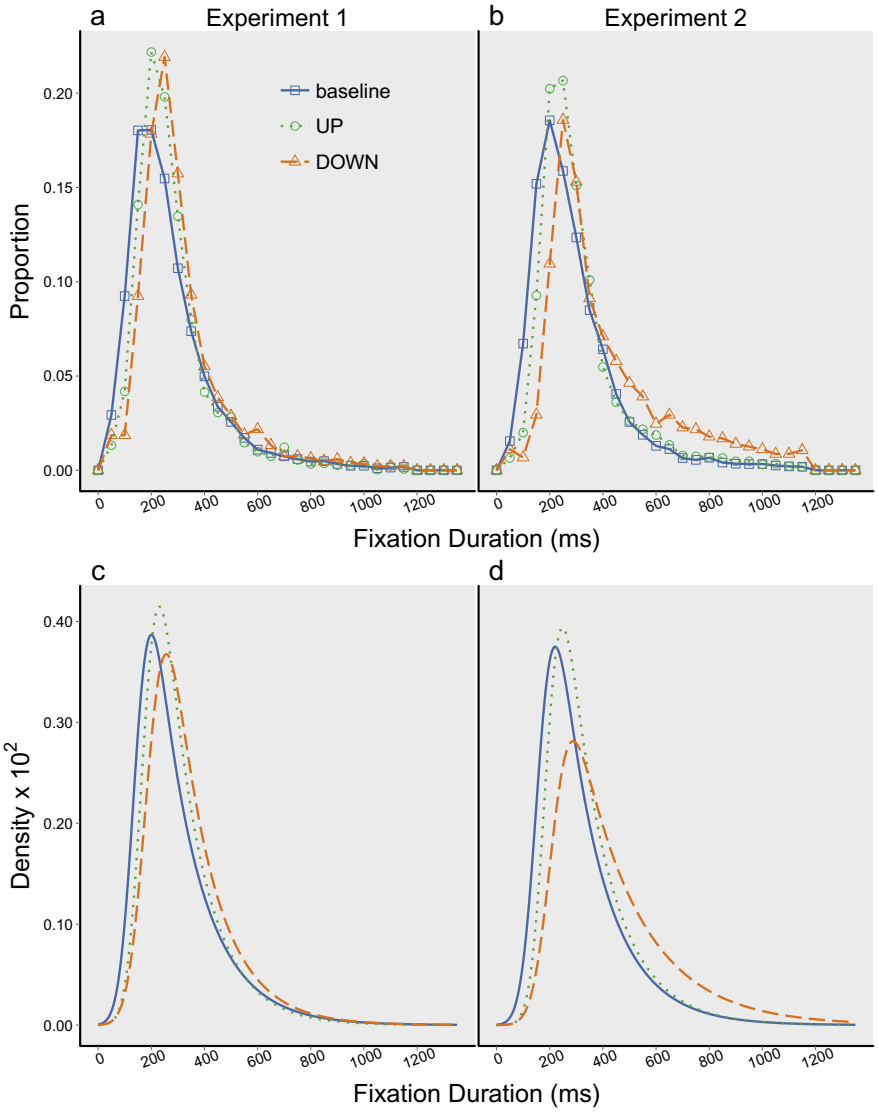

Fig. 3. Fixation duration distributions. Empirical distributions for the three luminance conditions in Experiments 1 (Panel a) and Experiment 2 (Panel b), and their respective ex-Gaussian fitted distributions plotted in (Panel c) and (Panel d).

a significant lengthening of fixation durations, but the magnitude of the increase to fixation durations was considerably smaller than in the DOWN condition. Taken together, the results are indicative of an asymmetrical pattern of control such that difficulties in scene processing are directly incorporated and result in longer fixation durations, whereas processing facilitation does not lead to a comparable decrease in fixation durations.

One possibility for the lack of a speedup in the UP condition is that the magnitude of the luminance difference between the baseline and increase in luminance was insufficient to provide enough processing facilitation to elicit shorter fixation durations. That is, the possibility remains that while a luminance shift from 80% to 60% is sufficient to create scene processing difficulties, a shift from 80% to 100% is insufficient to create a context for processing facilitation. This hypothesis is strengthened by the results of the distributional analyses. This analysis showed that in the UP condition, an overall shift in the distribution occurred due to a significant positive effect on μ. However, we also observed a significant negative influence on the tail of the distribution (decrease in τ), indicating a significantly less skewed distribution in the UP condition. Therefore, we hypothesise that a more extreme luminance enhancement may result in a diminished impact on the central tendency of the distribution than was observed in Experiment 1, but that the influence on the tail of the distribution will remain. Experiment 2 was designed to address this possibility by lowering the baseline luminance of the scene to 60% and further lowering the luminance to 20% in the DOWN condition and raising it to 100% in the UP condition.

2.3. Experiment 2

2.3.1. Methods

Procedure and stimuli. The procedure and stimuli for Experiment 2 were identical to that of Experiment 1 in all aspects other than the magnitude of the luminance change. During the saccade immediately preceding a critical fixation, the luminance was either shifted up to 100% or down to 20% luminance, from a 60% luminance baseline. During the saccade immediately following the critical fixation, the luminance of the scene was changed back to the 60% baseline level.

Participants. 13 females and 4 males who did not participate in Experiment 1 were tested in Experiment 2. The mean age of the participants was 24 years. Each participant was paid £7 per hour of participation in compensation for their time.

2.3.2. Results

Experiment 2 sought to complement the results observed in Experiment 1 by testing whether similar effects would be observed when a different baseline luminance level was used, and when the magnitude of the luminance shifts was increased. The observed pattern of mean durations is plotted in Fig. 2. In the LME model, the intercept for the fixed effect of luminance shift estimates the mean value for the no-shift condition ($b = 319.47, SE = 11.09, t = 28.79$). Experiment 2 used a lower baseline level of original scene luminance than Experiment 1 (60% vs. 80%). Accordingly, the mean fixation duration in the no-shift baseline luminance condition was longer in Experiment 2 than in Experiment 1 (319 ms vs. 297 ms, Fig. 2). Following the default prediction, downward luminance shifts were associated with critical fixations that were significantly longer than in the no-shift condition ($b = 124.28, SE = 13.15, t = 9.44$). Experiment 2 used a greater magnitude of luminance shifts than Experiment 1 (40% as opposed to 20%). Therefore, downward shifts in luminance resulted in a larger relative increase in fixation duration in Experiment 2 as compared to Experiment 1 (Fig. 2). In addition, there was again a significant increase in fixation durations for upward luminance shifts ($b = 24.55, SE = 7.92, t = 3.10$). Relative to the no-shift baseline condition, fixation durations increased by 38.9% in the DOWN condition but only 7.7% in the UP condition.

The approach to analysing the distributional effects in Experiment 2 was conducted along analogous lines to Experiment 1. Fig. 3b and d show the empirical and ex-Gaussian fitted distributions. A similar pattern was found to Experiment 1 in that the distributions showed a general rightward shift consistent with the increased mean durations observed in both luminance shift conditions. The GAMLSS model yielded a significant positive effect on μ for both the DOWN and UP conditions relative to the no-shift baseline condition (DOWN: $b = 51.63, SE = 3.65, t = 14.13$; UP: $b = 31.11, SE = 3.02, t = 10.28$). In the DOWN condition, there was a significant positive effect on σ ($b = 7.42, SE = 2.71, t = 2.73$), indicative of an increase in the variance in this condition. As in Experiment 1, there was no effect on σ in the UP condition ($b = -2.77, SE = 2.28, t = -1.21$). With regard to the τ parameter, a different pattern of results was observed than in Experiment 1. In the DOWN condition, there was a substantial increase in long fixation durations, which is manifested as a more positively skewed distribution. This late influence on the tail of the distribution was substantiated by a statistically significant positive effect of DOWN on the τ parameter ($b = 70.14, SE = 6.32, t = 11.08$). No statistically significant effect of the UP condition on τ was observed ($b = -6.51, SE = 4.62, t = -1.40$).

2.3.3. Discussion

A possible explanation for the observation of no facilitatory effect in Experiment 1 is that the magnitude of the luminance increase was too small to result in benefits in processing to the degree required in order to observe shortened fixation durations. Experiment 2 directly tested this hypothesis by increasing the magnitude of the luminance shift from baseline in both the UP and DOWN condition. Mean fixation durations observed in Experiment 2 showed a similar pattern to Experiment 1. Further decreasing the luminance of the scene during selected critical fixations was associated with an immediate and substantial increase in fixation duration. Furthermore, we did not observe a decrease in fixation durations following a facilitation in stimulus processing, as was assumed to occur following the increase in scene luminance. By inspecting the parameters of the ex-Gaussian distribution in Experiment 1, we speculated that if the more extreme luminance shift in the UP direction diminished the influence on the central tendency of the distribution then a facilitation effect may have been observed. The results from the analysis of means and parameter of the ex-Gaussian fit suggest that this is not the case. These results complement Experiment 1 and provide further support for the hypothesis that fixation durations are controlled in an asymmetric manner. The results from Experiment 1 and 2 show that a fixation-contingent increase of overall scene luminance was not sufficient to elicit a speedup in processing as observed through decreased fixation durations.

3. General discussion

Two experiments were conducted to test whether the adjustment of fixation durations in naturalistic scene viewing is unidirectional (slow down), or bidirectional (speed up and slow down). A saccade-contingent display change method was used to make the scene more difficult or easier to process during prespecified critical fixations. In Experiment 1, a luminance baseline of 80% was presented to participants and the luminance was shifted to either 60% (DOWN) or 100% (UP). Experiment 2 extended these results by reporting a similar pattern for a 60% baseline with shifts to 20% (DOWN) and 100% (UP). If the direct-control process was asymmetric or unidirectional, decreasing the luminance of the scene should make processing more difficult and result in longer fixations, while clarifying the scene by increasing the luminance should have no effect on the duration of critical fixations. In contrast, if fixation durations were controlled in a symmetric or bidirectional manner, shifting luminance down should result in longer fixation durations, and shifting luminance up should result in shorter fixation durations. In both experiments, a pattern consistent with the asymmetrical hypothesis was observed such that decreases to luminance resulted in longer fixation durations, but increases to luminance did not result in an immediate decrease in fixation durations.

Downward luminance shifts were associated with increases in fixation durations in both Experiment 1 and Experiment 2. This was reflected in a difference in elevated mean durations relative to the baseline luminance. The overall effect of decreasing luminance on fixation durations is broadly a replication of results reported by Henderson, Nuthmann, and Luke (2013) with different baseline conditions (60% and 80% compared to 100%) and novel stimuli. Additional distributional analyses using GAMLSS regression models qualified the time course of the observed effects. The results from the distributional analysis for Experiment 1 revealed that the increased durations in the DOWN condition occurred due to an overall shift in the distribution (increase in μ) as well as a significant increase in σ, the latter indicating the presence of greater variability in fixation durations in this condition. By comparison, the comparatively larger increase in durations in Experiment 2 was again associated with an overall shift in the distribution (increase in μ) and an increase in σ, but also with a longer tail (increase in τ). The specific influence on the tail of the distribution in Experiment 2 may be partially informed by a recent study conducted by Glaholt, Rayner, and Reingold (2013). In their study, the authors used a fixation-contingent scene quality paradigm to modify scenes under a variety of conditions such as spatial frequency filtering, and changes to the orientation of the image. In order to observe the differential effects of these modifications on fixation durations, they reported ex-Gaussian fitted distributions for the various conditions. They found that effects on the tail of the distributions were observed primarily for conditions in which the manipulation was hypothesised to result in a change that pre-

sented challenges to the later stages of stimulus encoding. In Experiment 2 of the current study, the extreme luminance manipulation (60% → 20%) is likely to lead to difficulties in integrating the lowered-luminance stimulus into existing working memory structures and may partially account for the overall increase in fixation durations and the effects observed on the tail of the distribution.

Upward luminance shifts were associated with a small but reliable increase in fixation durations, which is contrary to predictions by both the asymmetric control hypothesis (no change) and symmetric control hypothesis (decrease). The distributions revealed that in both experiments the increase was attributable to an increase in the central tendency (increased μ); there was no increase in τ in either experiment (rather a significant negative effect in Experiment 1), and no effect of UP on σ.

One explanation for this small increase comes from an effect of surprise that may accompany the shift of luminance that participants encounter on critical fixations. The analysis provided by Glaholt, Rayner, and Reingold (2013) is informative of why this might be the case. They found that fixation durations were increased in all conditions, but that the effects on the tail were absent for the conditions in which no encoding difficulty was to be expected. These contrasting effects were explained by suggesting that the fast-acting effect that influences the central tendency is a result of surprise due to a detected mismatch between transsaccadic stimulus content. The small but significant increase in fixation durations in the UP condition of both experiments reported here is consistent with the fast-acting effect of surprise that is hypothesised to occur following transsaccadic changes to the scene. Their study included another control experiment that is relevant to the interpretation of the present results. During critical fixations, colour information was added to the greyscale scene. By clarifying the stimulus with a colour enhancement, stimulus processing should be facilitated. According to the symmetric control hypothesis, adding colour should lead to an immediate decrease in fixation duration. However, an increase in the durations of critical fixations was observed, which resulted from an increase to μ, but not from τ. These results are consistent with the results reported here. Our presentation of the distributional effects that further qualify the inferences made by assessing differences in mean fixation durations is in keeping with recent analyses in reading (e.g., Glaholt, Rayner, & Reingold, 2014; Luke, Nuthmann, & Henderson, 2013; Reingold et al., 2012; Staub et al., 2010) and scene viewing (Glaholt & Reingold, 2012; Glaholt, Rayner, & Reingold, 2013; Luke, Nuthmann, & Henderson, 2013). Such analyses are highly informative in that they reveal the specific components of the distributions that contribute to the observed mean effects. As has been previously discussed, these results contribute to a growing body of research demonstrating consistent distributional effects within a variety of viewing tasks.

The pattern emerging from the present study, as well as recent empirical results, is that the direct control mechanism operates in an asymmetric manner, in both scene viewing and other visual-cognitive tasks. For instance, Glaholt, Rayner, and Reingold (2014) reported an asymmetrical control pattern in a reading task in which the contrast of the sentence text was either increased or decreased in a gaze-contingent manner. During the saccade immediately preceding a critical fixation the contrast of the sentence text with the background was either increased, decreased, or was left unchanged. The authors found that upon landing on a sentence that had decreased contrast, fixation durations were increased relative to the no change baseline condition, whereas fixation durations remained the same when contrast was increased. Such results complement previous results observed in both reading (Kennison & Clifton, 1995) and in visual search (Hooge, Vlaskamp, & Over, 2007).

The results reported here have direct theoretical consequences for models of eye-movement control generally, but most specifically for accounts of fixation behaviours in scene perception. A computational framework that has had considerable success in modelling the temporal aspects of eye-movement control in scene viewing is known as the CRISP model (Nuthmann et al., 2010; Nuthmann & Henderson, 2012). The CRISP model is a stochastic timing model such that a random walk timing process accumulates to a fixed threshold value. Once this threshold is reached, the programming of a saccade is initiated. The variability of fixation durations predicted from the model are generated from three primary sources, (a) the inherent stochasticity of the random walk timer, (b) modulation of the random walk's transition rate due to difficulties encountered during stimulus processing, and (c) cancellation and reprogramming of current saccade programmes. In the original formulation of the CRISP model it was assumed that eye-movement control operates in a manner consistent with what we have here called asymmetric control. That is, modulations to the timer could only occur due to processing difficulty that is expressed as a timer slowdown. With regards to the present results, the CRISP model captures such behaviour by assuming that difficulties in processing due to the decrease in luminance, result in a slowdown of the random walk timer rate and a temporal increase in the interval between successive saccade programmes. However, the results reported here with respect to the condition in which luminance is increased suggest that the default timer slowdown implemented in the CRISP model is sufficient to capture the effects of both degrading and enhancing stimulus processing.

A relevant question for future studies is how the adaptation of fixation durations to immediate changes in processing difficulty changes over the course of viewing. One possibility is that fixation durations may adapt with an immediate increase when processing difficulty increases but may decrease more gradually, say on the second or third fixation, following a decrease in difficulty. Trukenbrod and Engbert (2014) have investigated this issue using a task that required participants to scan sequences of horizontally arranged symbols from left to the right to search for a target stimulus. The target was always a ring, and Landolt-Cs were used as distractors. Processing difficulty of the stimulus elements was manipulated by increasing or decreasing the size of the gap in the Landolt-Cs. Fixation durations upon first encountering a change of difficulty, as well as fixation times on subsequent elements were measured. The authors reported an asymmetrical pattern of control of fixation durations such that increasing difficulty resulted in an increase to fixation durations upon first encountering the change, while a decrease in difficulty resulted in no immediate impact. However, they reported a delayed adjustment to fixation durations in the decreasing difficulty condition, as fixation durations showed evidence of a decrease for later fixations. The time course of the adjustment to changes in processing difficulty within scene viewing is currently an open empirical question.

4. Conclusion

In summary, this study used a luminance manipulation in order to vary the scene processing difficulty in a gaze-contingent fashion on critical fixations. We predicted that if the control of fixation durations operates in a symmetric manner, then shifting luminance down would result in increased fixation durations, while shifting luminance up would result in decreased fixation durations. On the other hand, if control is asymmetric we predicted that decreasing the luminance would result in fixation duration increases and that luminance increases would result in no change to fixation durations. The pattern of results observed in the two

experiments provides support for the asymmetric control hypothesis.

Acknowledgment

The authors would like to thank Alicia Forsberg for assistance with data collection. Portions of this research were presented at the 17th European Conference on Eye Movements (Lund, Sweden, August 2013).

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes (R package version 0.999999-0).

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research, 19*(9), 967–983.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behaviour Research Methods Instruments & Computers, 34*(4), 613–617.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777–813.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2013). Spatial frequency filtering and the direct control of fixation durations during scene viewing. *Attention, Perception, & Psychophysics, 75*(8), 1761–1773.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2014). A rapid and immediate effect of stimulus quality on the duration of individual fixations during reading. *Visual Cognition, 22*(3), 377–389.

Glaholt, M. G., & Reingold, E. M. (2012). Direct control of fixation times in scene viewing: Evidence from analysis of the distribution of first fixation duration. *Visual Cognition, 20*(6), 605–626.

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 318–322.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*(3), 566–573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17*(6–7), 1055–1082.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research, 38*(9), 1295–1302.

Hooge, I. T. C., Vlaskamp, B. N. S., & Over, E. A. B. (2007). Saccadic search: On the duration of a fixation. In R. P. G. Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 581–595). Elsevier.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier.

Kennison, S. M., & Clifton, C. Jr., (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: General, 21*(1), 68–81.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 14.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General, 114*(3), 342–356.

Luke, S. G., Nuthmann, A., & Henderson, J. M. (2013). Eye movement control in scene viewing and reading: Evidence from the stimulus onset delay paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 10–15.

McConkie, G. W., & Loschky, L. C. (2002). Perception onset time during fixations in free viewing. *Behaviour Research Methods Instruments & Computers, 34*(4), 481–490.

Murray, W. S., Fischer, M. H., & Tatler, B. W. (2013). Serial and parallel processes in eye movement control: Current controversies and future directions. *Quarterly Journal of Experimental Psychology, 66*(3), 417–428.

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition, 20*(4–5), 457–494.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*(2), 382–405.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210.

Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics, 73*(4), 1120–1132.

R Core Team (2012). R: A language and environment for statistical computing. R foundation for statistical computing.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology, 33A*, 351–373.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review, 119*(1), 155–185.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65*(2), 177–206.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C-Applied Statistics, 54*(3), 507–544.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113–121.

SR Research Ltd. (2006). EyeLink programmer's guide (3.0 ed.).

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1280–1293.

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*. http://dx.doi.org/10.3758/s13423-013-0575-0.

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). Visual memory for fixated regions of natural images dissociates attraction and recognition. *Perception, 38*(8), 1152–1171.

Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research, 46*(3), 417–425.

Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: The effect of element spacing. *Experimental Brain Research, 167*(2), 246–259.

Walshe, R. C., & Nuthmann, A. (2013). Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1569–1574). Berlin, Germany: Cognitive Science Society.

BIBLIOGRAPHY

Aslin, R. N., & Shea, S. L. (1987). The amplitude and angle of saccades to double-step target displacements. *Vision Research*, *27*, 1925–1942.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D. M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes (R package version 0.999999-0)*.

Becker, W. (1991). Saccades. In R. Carpenter (Ed.), *Eye-Movements: Vol. 8 in Vision and visual Dysfunction* (pp. 95—137). London: Macmillan.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*, 967–983.

Berniker, M., & Kording, K. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, *11*, 1454–1461.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Burr, D., & Cicchini, G. M. (2014). Vision: Efficient Adaptive Coding. *Current Biology*, *24*, R1096–R1098.

Camalier, C., Gotler, A., Murthy, A., Thompson, K., Logan, G., Palmeri, T., & Schall, J. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*, 2187–2211.

Carpenter, R., & Williams, M. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*, 59–62.

Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Pro-*

ceedings of the National Academy of Sciences of the United States of America, *101*, 13086–13090.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behaviour Research Methods Instruments & Computers*, *34*, 613–617.

Deep, K., Singh, K. P., Kansal, M., & Mohan, C. (2009). A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation*, *212*, 505–518.

DeGroot, M., & Schervish, M. (2012). *Probability and Statistics*. (4th ed.). Pearson.

Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, *6*, 159–178.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Einhäuser, W., Rutishauser, U., Frady, E. P., Nadler, S., König, P., & Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. *Journal of Vision*, *6*, 1148–1158.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*, 1–19.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*, 621–636.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813.

Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. In A. G. Gale, & F. John-

son (Eds.), *Theoretical and Applied Aspects of Eye Movement Research*. Amsterdamn: North Holland.

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*, 1777–1790.

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, *4*, 563–572.

Gillespie, D. T. (1978). Monte carlo simulation of random walks with residence time dependent transition probability rates. *Journal of Computational Physics*, *28*, 395–407.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2013). Spatial frequency filtering and the direct control of fixation durations during scene viewing. *Attention, Perception, & Psychophysics*, *75*, 1761–1773.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2014). A rapid effect of stimulus quality on the durations of individual fixations during reading. *Visual Cognition*, *22*, 377–389.

Glaholt, M. G., & Reingold, E. M. (2012). Direct control of fixation times in scene viewing: Evidence from analysis of the distribution of first fixation duration. *Visual Cognition*, *20*, 605–626.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital Image Processing Using MATLAB*. Knoxville, TN: Gatesmark Publishing.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics*, *59*, 323–346.

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 318–322.

Henderson, J. M., Olejarczyk, J., Luke, S. G., & Schmidt, J. (2014). Eye movement control during scene viewing: Immediate degrada-

tion and enhancement effects of spatial frequency filtering. *Visual Cognition: Special Issue Honoring George W. McConkie*, *22*, 486–502.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, *15*, 566–573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, *17*, 1055–1082.

Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*, 204–210.

Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, *12*, 1–13.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, *38*, 1295–1302.

Hooge, I. T. C., Vlaskamp, B. N. S., & Over, E. A. B. (2007). Saccadic search: On the duration of a fixation. In R. P. G. Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 581–595). Oxford: Elsevier.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*, 71–77.

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 580–593.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier.

Kaspar, K., & König, P. (2011). Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. *PLoS One*, *6*, e21719.

Kennison, S. M., & Clifton, C., Jr (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: General*, *21*, 68–81.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*, 14.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565.

Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, *13*, 11.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological review*, *104*, 524.

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). Visual memory for fixated regions of natural images dissociates attraction and recognition. *Perception*, *38*, 1152–1171.

Loftus, G. R. (1985). Picture perception: effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, *114*, 342–356.

Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, *97*, 795–805.

Luke, S. G., Nuthmann, A., & Henderson, J. M. (2013). Eye Movement control in scene viewing and reading: Evidence from the stimulus

onset delay paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 10–15.

McConkie, G. W., & Loschky, L. C. (2002). Perception onset time during fixations in free viewing. *Behaviour Research Methods Instruments & Computers*, *34*, 481–490.

McDonald, S. A., Carpenter, R. H. S., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review*, *112*, 814–840.

McPeek, R., Skavenski, A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision research*, *40*, 2499–2516.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*, 1–15.

Murray, W. S., Fischer, M. H., & Tatler, B. W. (2013). Serial and parallel processes in eye movement control: Current controversies and future directions. *Quarterly Journal of Experimental Psychology*, *66*, 417–428.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 342–360.

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, *1339*, 82–86.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*, 322–336.

Nuthmann, A., & Henderson, J. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, *20*, 457–494.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*, 1–19.

Nuthmann, A., & Kliegl, R. (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision*, *29*, 1–28.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*, 382–405.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, *73*, 1120–1132.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. R Foundation for Statistical Computing Vienna, Austria. R package version 3.1-117.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.

Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*, 721–763.

Ray, S., Schall, J. D., & Murthy, A. (2004). Programming of double-step saccade sequences: modulation by cognitive control. *Vision Research*, *44*, 2707–2718.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–506.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *The Quarterly Journal of Experimental Psychology*, *33*, 351–373.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review*, *119*, 155–185.

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, *7*, 1–6.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*, 177–206.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C-Applied Statistics*, *54*, 507–544.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*, 113–121.

Saslow, M. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1024–1029.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, *20*, 391–421.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, *2*, 33–42.

Shen, K., & Paré, M. (2014). Predictive saccade target selection in superior colliculus during visual search. *The Journal of Neuroscience*, *34*, 5640–5648.

Shevell, F. A. A., S. K. amd Kingdom (2008). Color in complex scenes. *Annual Review of Psychology*, *59*, 143–166.

Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, *53*, 305–314.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*, 1–16.

SR Research Ltd. (2006). *EyeLink Programmer's Guide* (3rd ed.).

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1280–1293.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*, 1–18.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review.*, *21*, 904–34.

Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research*, *46*, 417–425.

Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: the effect of element spacing. *Experimental Brain Research*, *167*, 246–259.

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and Bayesian inference: A study of saccadic response speed. *Cerebral Cortex*, *24*, 1436–1450.

Wallman, J., & Fuchs, A. F. (1998). Saccadic gain modification: Visual error drives motor adaptation. *Journal of Neurophysiology*, *80*, 2405–2416.

Walshe, R. C., & Nuthmann, A. (2013). Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1569–1574). Berlin, Germany: Cognitive Science Society.

Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, *100*, 38–46.

Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision*, *15(5)*, 1–19.

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, *52*, 932–941.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology*, *9*.

Wu, E. X. W., Gilani, S. O., van Boxtel, J. J. A., Amihai, I., Chua, F. K., & Yen, S.-C. (2013). Parallel programming of saccades during natural scene viewing: evidence from eye movement positions. *Journal of Vision*, *13*, 1–14.